



UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO DE CIÊNCIAS EXATAS, NATURAIS E DA SAÚDE - CCENS
PROGRAMA DE PÓS – GRADUAÇÃO EM AGROQUÍMICA



CAMILA FERNANDES VIEIRA

**AVALIAÇÃO DA ESPECTROMETRIA NO INFRAVERMELHO PRÓXIMO NA
DETERMINAÇÃO DE C E N DE SOLOS DA REGIÃO SUL DO ESPÍRITO SANTO**

ALEGRE-ES

2019

CAMILA FERNANDES VIEIRA

**AVALIAÇÃO DA ESPECTROMETRIA NO INFRAVERMELHO PRÓXIMO NA
DETERMINAÇÃO DE C E N DE SOLOS DA REGIÃO SUL DO ESPÍRITO SANTO**

Dissertação apresentada ao programa de Pós-Graduação em Agroquímica do Centro de Ciências Exatas, Naturais e da Saúde da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do título de Mestre em Agroquímica, na área de concentração Química Ambiental.

Orientador: Prof. Eduardo de Sá Mendonça

Coorientador: Prof. Diego Lang Burak

ALEGRE-ES

2019

Ficha catalográfica disponibilizada pelo Sistema Integrado de
Bibliotecas - SIBI/UFES e elaborada pelo autor

V657a Vieira, Camila Fernandes, 1991-
Avaliação da espectrometria no infravermelho próximo na
determinação de C e N de solos da região Sul do Espírito Santo /
Camila Fernandes Vieira. - 2019.
60 f.

Orientador: Eduardo de Sá Mendonça.

Coorientadores: Diego Lang Burak, André Marcelo de
Souza.

Dissertação (Mestrado em Agroquímica) - Universidade
Federal do Espírito Santo, Centro de Ciências Exatas, Naturais e
da Saúde.

1. Mínimos quadrados parciais. 2. Calibração multivariada. 3.
Espectrometria de Infravermelho próximo. I. Mendonça, Eduardo
de Sá. II. Burak, Diego Lang. III. Souza, André Marcelo de.
IV. Universidade Federal do Espírito Santo. Centro de Ciências
Exatas, Naturais e da Saúde. V. Título.

CDU: 631.41

CAMILA FERNANDES VIEIRA

**AVALIAÇÃO DA ESPECTROMETRIA NO INFRAVERMELHO PRÓXIMO NA
DETERMINAÇÃO DE C E N DE SOLOS DA REGIÃO SUL DO ESPÍRITO SANTO**

Dissertação apresentada ao programa de Pós-Graduação em Agroquímica do Centro de Ciências Exatas, Naturais e da Saúde da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do título de Mestre em Agroquímica, na área de concentração Química Ambiental.

Aprovado em 15 de Maio de 2019.

COMISSÃO EXAMINADORA



Prof. Dr. Eduardo de Sá Mendonça
Universidade Federal do Espírito Santo
Orientador (a)



Prof. Dr. Diego Lang Burak
Universidade Federal do Espírito Santo
Coorientador



Prof. Dr. Alexandre dos Santos Anastácio
Universidade Federal do Espírito Santo
Examinador

À minha mãe e minha família, por confiarem e acreditarem na minha capacidade de chegar até aqui.

AGRADECIMENTOS

Aos meus pais, Margareth e Aureo, que me apoiaram todo esse tempo em Alegre, me ensinando a ser forte e não desistir jamais dos objetivos que eu queria alcançar. Eles confiaram em mim e acreditaram no meu potencial de chegar até aqui.

A minha avó Enedina que sempre em suas orações pede pela minha saúde e pede proteção a mim por estar longe de casa.

Ao meu orientador Eduardo de Sá Mendonça, pela confiança no meu trabalho e que nos meus piores momentos me deu a mão e me deu força para continuar e não desistir da minha caminhada.

Ao meu coorientador Diego Lang Burak, pelas cobranças, auxílio e sempre estar disponível.

Ao professor Adilson Vidal Costa, que sempre esteve me ajudando e me aconselhando em continuar e também a não desistir de chegar ao fim e virar mestre.

Aos professores Alexandre dos Santos Anastácio e Demetrius Profeti, que estiveram ao meu lado me auxiliando e proporcionando todo apoio necessário. Agradeço por terem feito parte dessa trajetória.

Aos amigos do laboratório de rotina de solos do CCAE UFES, Yuri, Ana Paula, Ana Carla, Allison, Ravínia, que em todo o processo estiveram me ajudando nas análises.

Ao amigo Ramon Azevedo Braz, que sem ele o NIR não teria resultados.

Ao Udson Oliveira Ramos de Aguiar que especialmente esteve comigo aturando meus desesperos, minhas correrias e meus esquecimentos, na maior parte crítica do projeto.

Ao laboratório de rotina CCAE UFES, pelo fornecimento das amostras de solos.

A todas as pessoas que direta ou indiretamente ajudaram no processo construtivo desse trabalho.

“Talvez não tenha conseguido fazer o melhor, mas lutei para que o melhor fosse feito. Não sou o que deveria ser, mas Graças a Deus, não sou o que era antes”.

Marthin Luther King

RESUMO

VIEIRA, CAMILA FERNANDES. **Avaliação da espectrometria no infravermelho próximo na determinação de C e N de solos da região Sul do Espírito Santo.** 2019. 60p. Dissertação (Mestrado em Agroquímica) – Centro de Ciências Exatas, Naturais e da Saúde - CCENS, Universidade Federal do Espírito Santo, Alegre, ES. Orientador: Eduardo de Sá Mendonça. Coorientador: Diego Lang Burak.

O presente estudo teve como objetivo investigar a capacidade de predição dos teores totais de carbono orgânico e nitrogênio, por meio da espectroscopia de reflectância no infravermelho próximo (NIR) e o seu potencial para o uso em análises rotineiras de solos. Tal método apresenta maior agilidade na busca de resultados e abdica o uso de reagentes químicos danosos ao ambiente e a saúde humana. O trabalho constou das seguintes etapas: preparo de amostras de solo (TFSA – Terra fina seca ao ar – peneiradas em 0,2 mm); determinação dos teores de carbono orgânico total e nitrogênio total, seleção dos grupos de amostras para calibração e validação, aplicação de pré-tratamentos nos espectros e ajustes de modelos de mínimos quadrados parciais. Os métodos de pré-processamentos aplicados foram: Primeira e segunda derivada; remoção contínua; variação normal padrão; Savitzky-Golay; Savitzky-Golay com primeira derivada; correção multiplicativa de sinal. As amostras de carbono orgânico total e nitrogênio total selecionadas para calibração e validação, obtiveram estatística descritiva semelhante. Os melhores ajustes do modelo em relação aos valores observados foram: teor de carbono orgânico total: R^2 – segunda derivada (0,87 e 0,89), para calibração e para validação respectivamente; teor de nitrogênio total: R^2 para calibração – variação normal padrão (0,68) e para validação – Savitzky-Golay (0,66). Ao avaliar os valores dos erros quadráticos médios entre os dados de calibração e validação, percebe-se que são estatisticamente semelhantes para os dois atributos em análise. O método espectroscopia NIR neste trabalho apresentou valores satisfatórios e com bons ajustes para calibração e validação tanto para análise de COT, quanto para análise de NT.

Palavras chave: Mínimos Quadrados parciais. Calibração multivariada. Espectrometria de Infravermelho próximo.

ABSTRACT

VIEIRA, CAMILA FERNANDES. **Evaluation of the near infrared spectrometry in the determination of C and N of soils of the southern region of Espírito Santo.** 2019. 60p. Dissertation (Master in Agrochemistry) – Center for Exact, Natural and Health Sciences - CCENS, Federal University of Espírito Santo, Alegre, ES. Adviser: Eduardo de Sá Mendonça. Co-adviser: Diego Lang Burak.

The present study aimed to investigate the predictive capacity of total organic carbon and nitrogen contents by means of near infrared reflectance spectroscopy (NIR) and its potential for use in routine soil analysis. This method presents greater agility in the search for results and abdicates the use of chemical reagents harmful to the environment and human health. The work consisted of the following steps: preparation of soil samples (TFSA - Dry earth in air - sieved in 0.2 mm); determination of the total organic carbon and total nitrogen contents, selection of groups of samples for calibration and validation, application of pre-treatments in the spectra and adjustments of partial least squares models. The pre-processing methods applied were: First and second derivative; continuous removal; standard normal variation; Savitzky-Golay; Savitzky-Golay with first derivative; Multiplicative signal correction. Samples of total organic carbon and total nitrogen selected for calibration and validation, obtained similar descriptive statistics. The best adjustments of the model in relation to the observed values were: total organic carbon content: R² - second derivative (0.87 and 0.89), for calibration and for validation respectively; total nitrogen content: R² for calibration - standard normal variation (0.68) and for validation - Savitzky-Golay (0.66). When evaluating the mean square error values between the calibration and validation data, it is noticed that they are statistically similar for the two attributes under analysis. The NIR spectroscopy method in this work presented satisfactory values and with good adjustments for calibration and validation for both TOC analysis and NT analysis.

Keywords: Partial least squares. Multivariate calibration. Near Infrared Spectrometry.

LISTA DE FIGURAS

Figura	Página
Figura 1: Ilustração do procedimento de calibração multivariada	20
Figura 2: Gráfico que ilustra o comportamento da primeira derivada	25
Figura 3: - Gráfico de um espectro contendo declive e gráfico de um espectro de segunda derivada onde o declive foi eliminado	26
Figura 4: Processo geral de construção de modelos de regressão	27
Figura 5: Estatística descritiva dos conjuntos de calibração e validação para COT. (A1) e (A2) densidade populacional definido pela média e desvio padrão. (B1) e (B2) limite normal esperado para o número de observação pelo gráfico de curtose. (C1) e (C2) gráfico de dispersão	37
Figura 6: Estatística descritiva dos conjuntos de calibração e validação para NT. (A1) e (A2) densidade populacional definido pela média e desvio padrão. (B1) e (B2) limite normal esperado para o número de observação pelo gráfico de curtose. (C1) e (C2) gráfico de dispersão	39
Figura 7: Reflectância dos espectros na região NIR ($4000\text{ cm}^{-1} - 10000\text{cm}^{-1}$) das amostras utilizadas para calibrar e validar os modelos	40
Figura 8: Reflectância dos espectros na região NIR ($4000\text{ cm}^{-1} - 10000\text{cm}^{-1}$) das amostras com teor mínimo e máximo de COT	41
Figura 9: Reflectância dos espectros na região NIR ($4000\text{ cm}^{-1} - 10000\text{cm}^{-1}$) das amostras com teor mínimo e máximo de NT	41
Figura 10: Reflectância dos espectros na região NIR ($4000\text{ cm}^{-1} - 10000\text{cm}^{-1}$) sem transformação (ST)	44

Figura 11: Reflectância dos espectros na região NIR pré processados por (A) primeira derivada, (B) segunda derivada, (C) remoção contínua, (D) variação normal padrão, (E) Savitzky-Golay e (F) Savitzky-Golay com primeira derivada	45
Figura 12: Desempenho do modelo para calibração dos dados de COT	49
Figura 13: Desempenho do modelo para calibração dos dados de NT	50
Figura 14: Desempenho do modelo para validação dos dados de COT	51
Figura 15: Desempenho do modelo para validação dos dados de NT	51
Figura 16: Coeficientes da equação em cada número de onda para análise de COT	52
Figura 17: Coeficientes da equação em cada número de onda para análise de NT	52

LISTA DE TABELAS

Tabela		Pág
Tabela 1-	Resumo dos trabalhos relacionados a técnica NIR e análise de solos	19
Tabela 2-	Estatísticas descritiva dos teores de COT (g Kg^{-1}) avaliadas em laboratório de rotina para o conjunto de calibração e validação.....	36
Tabela 3-	Estatísticas descritiva dos teores de NT (g Kg^{-1}) avaliadas em laboratório de rotina para o conjunto de calibração e validação	38
Tabela 4-	Avaliação dos modelos de calibração e validação obtidos por mínimos quadrados parciais na determinação de COT	46
Tabela 5-	Avaliação dos modelos de calibração e validação obtidos por mínimos quadrados parciais na determinação de NT	47

SUMÁRIO

1	INTRODUÇÃO	11
2	REFERÊNCIAL TEÓRICO	13
2.1	CARBONO ORGÂNICO DO SOLO	13
2.2	NITROGÊNIO TOTAL DO SOLO	15
2.3	ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO (NIR)	16
2.3.1	Princípios da espectroscopia NIR	17
2.3.2	Aplicações da espectroscopia NIR	18
2.4	ANÁLISE MULTIVARIADA	20
2.4.1	Calibração Multivariada	21
2.4.1.1	Regressão por quadrados mínimos parciais (PLS)	23
2.4.1.2	Validação do modelo	28
3	MATERIAL E MÉTODOS	29
3.1	CARBONO ORGÂNICO TOTAL DO SOLO	29
3.2	NITROGÊNIO TOTAL DO SOLO	30
3.3	ESPECTROSCOPIA DE INFRAVERMELHO PRÓXIMO - EQUIPAMENTO NIR	31
3.4	AVALIAÇÃO DE OUTLIERS	32
3.5	SELEÇÃO CONJUNTO DE CALIBRAÇÃO E VALIDAÇÃO	32
3.5.1	Carbono Orgânico Total	32
3.5.2	Nitrogênio Total	32
3.6	PRÉ-TRATAMENTOS DOS ESPECTROS	33
3.6.1	Ajuste do modelo de mínimos quadrado parciais	33
4	RESULTADOS E DISCUSSÃO	36
4.1	ESTATÍSTICA DESCRITIVA	36
4.1.1	Carbono Orgânico Total	36
4.1.2	Nitrogênio Total	38
4.2	ESPECTROS NO NIR E CARACTERÍSTICAS DAS CURVAS DE REFLECTÂNCIA ESPECTRAL DO SOLO	39
4.3	PRÉ PROCESSAMENTO DOS DADOS	43
4.3.1	Calibração Multivariada - PLS	46
5	CONCLUSÕES	53
6	REFERÊNCIAS	54

1 INTRODUÇÃO

O solo é uma parte imprescindível quando se trata da produção vegetal e toda a sua cadeia produtiva. Atributos químicos e físicos do solo controlam sua qualidade e condições edáficas, quesitos importantes na produtividade das culturas. O estudo de tais atributos no contexto da fertilidade do solo garante ao produtor e profissional da área informações essenciais para o manejo do solo e busca de resultados satisfatórios dentro da agricultura. Um dos principais fatores a ser observado quando se trata da qualidade e condição do solo é o teor de matéria orgânica do solo (MOS), já que a mesma desempenha função importante em relação à oferta de nutrientes, propriedades químicas e físicas do solo (CIOTTA et al., 2003), e é a principal responsável pela retenção dos nutrientes, especialmente em solos de textura arenosa (SANTOS; PEREIRA; KORNDÖRFER, 2010). Uma característica que faz a matéria orgânica (MO) ser um dos atributos importantes de serem analisados é por ter a função de ciclagem de nutrientes e ser a maior reserva terrestre de C dentro do seu ciclo global (FRANCESCHINI et al., 2013). Os teores de N, nas diferentes profundidades do solo, apresentam o mesmo padrão de distribuição dos teores de C, com os maiores valores observados em camadas superficiais do solo, onde há maior acúmulo de MO (RANGEL et al., 2008).

A técnica mais utilizada de determinação do C do solo é a oxidação por dicromato (Método via úmida – Walkley-Black, 1934), considerada técnica que demanda maior custo e gera resíduos ambientalmente nocivos ao ambiente, como o cromo. Essas análises de rotina demandam tempo e reagentes. E é nesse contexto que pesquisas na busca por alternativas aos processos laboratoriais tradicionais demonstram a eficácia do uso da espectroscopia de infravermelho próximo (NIR) em análises de solos, principalmente na predição de parâmetros como teor de carbono orgânico, nitrogênio total, capacidade de troca iônica, entre outros (CHANG et al, 2001; SHEPHERD; WALSH, 2002). A técnica de NIR se fundamenta na medida da intensidade de absorção de radiação eletromagnética na região do NIR (750-2500 nm) medida pelos detectores, e depende da natureza das moléculas e da sua concentração que absorvem a radiação eletromagnética em diferentes comprimentos de onda. Os resultados da refletância difusa são relacionados a atributos do solo para

geração dos modelos e calibração dos métodos podendo vir a substituir grande parte dos métodos tradicionais, garantindo a qualidade e a precisão presente em cada análise (SANTOS; PEREIRA; KORNDÖRFER, 2010). Essa técnica analítica possui a vantagem de ser simples, rápida, não destrutiva, não necessita de pré-tratamento em laboratório, além de não ser poluente ao meio ambiente, pois não utiliza-se reagentes químicos contaminantes. Com isso, o NIR vem alcançando todos os objetivos citados e vem sendo utilizada cada vez mais como ferramenta analítica (FIDENCIO, 2001). O carbono orgânico total têm sido avaliados pelo NIR de forma satisfatória. Enquanto que, atributos químicos como capacidade de troca catiônica (CTC), teores de nutrientes têm sido avaliados pelo NIR com respostas contrastantes entre os estudos (CHANG et al, 2001; SHEPHERD; WALSH, 2002; VISCARRA ROSSEL et al., 2009; MORÓN e COZZOLINO, 2003).

Com a hipótese de que pode-se aperfeiçoar a obtenção de resultados de análises de C e N em amostras de solos, sem diminuir a precisão e exatidão a partir do NIR, o presente trabalho tem como objetivo investigar a predição dos teores de C e N, obtidos em laboratórios, a partir da espectrometria do infravermelho próximo de um extenso banco de dados de solos da região Sul do Estado do Espírito Santo.

2 REFERÊNCIAL TEÓRICO

2.1 CARBONO ORGÂNICO DO SOLO

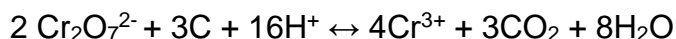
Dentre os constituintes mais importantes do solo a matéria orgânica ou carbono orgânico é o componente que está presente em menor quantidade em comparação à fase mineral mas é a principal fonte reservatória de energia e de nutrientes que estarão prontamente disponíveis para serem absorvidos pelas plantas ao interagirem com a matéria orgânica por meio de ligações eletrostáticas. Normalmente, as frações orgânicas estão compostas entre 1 e 5% do solo, com exceção os Organossolos que contém mais de 8% de carbono orgânico. Essa variação da sua quantidade no solo possui grande influência tanto nas propriedades químicas, tais como: pH, CTC, na capacidade de suprir o nitrogênio, fósforo e enxofre, quanto nas propriedades físicas: tais como: cor do solo, agregação com implicações na densidade, porosidade, aeração, capacidade de retenção de água e teor de argila (FIDÊNCIO, 2001; ROMÃO, 2012; LEITE, 2004). A quantificação da matéria orgânica do solos é feita, em laboratório, a partir do conteúdo de carbono orgânico. A fórmula empregada é: matéria orgânica (g Kg^{-1}) = $1,724 \times \text{C}$ (fator de Van Bemmelen), no qual baseia-se que a matéria orgânica possui 58% de carbono orgânico (SANTOS A. P., 2011). O teor de carbono orgânico no solo é um dos índices para diferenciar solos minerais de solos orgânicos e determinar a dose de N para as culturas agrícolas (SANTOS A. P., 2011). Em solos brasileiros, a carbono orgânico pode contribuir com até 80% das cargas negativas do solo.

O carbono orgânico do solo pode ser determinado por oxidação seca ou úmida. O método de maior exatidão é pela oxidação seca que mede-se o CO_2 produzido pela combustão do solo, mas em contrapartida ele requer equipamentos de custo elevado. A oxidação por via úmida é a mais utilizada devido ao baixo custo, na qual o C do solo é oxidado por uma solução oxidante. Assume-se que todo C do solo esteja em um estado de oxidação zero (C), o que na prática, não é verdade (MENDONÇA & MATOS, 2017). O método mais empregado é o proposto por Walkley-Black. O método utiliza o dicromato ($\text{Cr}_2\text{O}_7^{2-}$) (Cr VI) em meio ácido (H_2SO_4) como agente oxidante com auxílio de ácido fosfórico concentrado para facilitar a visualização do ponto final da titulação,

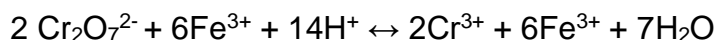
tendo como titulante uma solução de sulfato ferroso amoniacal. O ácido fosfórico impede que a coloração amarelada, obtida da oxidação de Fe^{2+} a Fe^{3+} na titulação, dificulte a visualização do ponto final da titulação, quando a solução assume a cor verde proveniente da redução do Cr^{6+} a Cr^{3+} , como mostra a equação (FIDÊNCIO, 2001).



Considera-se que a reação de oxidação seja:



Porém, somente a utilização de dicromato em meio ácido não é suficiente para oxidar todo o C, o que se torna necessário a utilização de uma fonte externa de calor para maximizar a oxidação do C pelo dicromato, assim certifica-se que seja oxidado. A dosagem pode ser feita por meio da determinação colorimétrica do Cr^{3+} formado, ou por meio da titulação do dicromato (Cr^{6+}) restante da oxidação, quando este é colocado em excesso. A titulação do dicromato é feita com uma solução de ferro reduzido ($\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$) em meio ácido, empregando-se como indicador o ferroin, conforme a reação (MENDONÇA & MATOS, 2017):



Alguns problemas associados aos métodos que utilizam dicromato envolvem suposições sobre o estado de oxidação do COS, pois assumem que o CO tem estado de oxidação zero e peso equivalente a 3g por equivalente quando reage com o dicromato. Na metodologia de oxidação do COT por íons dicromato são gerados resíduos de Cr (VI), este apresenta alta toxicidade, pois tem alto poder mutagênico que pode causar câncer. Apesar da reação do procedimento, contemplar a redução do Cr(VI) a Cr(III), ainda resta material sem reagir, pois é necessário a adição de uma grande quantidade de dicromato. O resíduo gerado exige um trabalho adicional do tratamento do resíduo tóxico, antes do descarte (BENACI, 2010). Como ressalvas destacam-se a esse método: as incertezas associadas às etapas de preparo e titulação; o elevado consumo de reagentes químicos como, por exemplo, dicromato de potássio, ácido sulfúrico e sulfato ferroso amoniacal; o volume significativo de resíduos agressivos ao meio ambiente e saúde humana; e o elevado tempo da determinação desse atributo do solo.

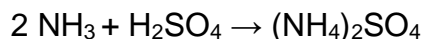
2.2 NITROGÊNIO TOTAL DO SOLO

O N converte-se rapidamente da forma amoniacal para nitrato pode se apresentar tanto na forma catiônica, quanto na forma aniônica (amoniacal), sendo ambas disponíveis à nutrição das plantas, é um dos nutrientes mais exigidos pelas plantas. Devido à sua alta exigência, é caracterizado como um dos principais fatores limitantes ao crescimento e desenvolvimento das plantas, que com sua deficiência apresentam-se amareladas e com crescimento reduzido (FIDÊNCIO, 2001; MENECHIN et al, 2008). Nos solos, o N ocorre tanto na forma orgânica quanto na inorgânica, e é fixado pela água da chuva por motivo das descargas elétricas de tempestades, e também através das bactérias fixadoras de N_2 de vida livre ou que realizam simbiose com as plantas. A matéria orgânica é a principal fonte de N no solo, mas no entanto, a maior parte desse N não está rapidamente disponível para as plantas e uma maneira de ganhar esse nutriente de forma mais ágil e com maior facilidade é com a aplicação de fertilizantes contendo a forma molecular (N_2 , que é inerte) ou em forma de nitrato (NO_3^-) (FIDÊNCIO, 2001; MENECHIN et al, 2008).

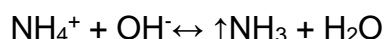
Os teores de N total dos solos agrícolas variam entre 0,5 e 5 g Kg^{-1} sendo que 98% destes encontram-se na forma orgânica. Os 2% restantes correspondem a formas inorgânicas, principalmente NH_4^+ e NO_3^- (MENDONÇA & MATOS, 2017). Vários fatores influenciam as formas da disponibilidade de N, tais como: temperatura, umidade, aeração, ciclos de umedecimento e secagem do solo, tipo de material orgânico, microrganismos, pH, preparo e fertilidade do solo, etc (FIDÊNCIO, 2001). Para avaliar os teores de N total no solo, alguns autores têm mostrado resultados favoráveis para o uso de extratores químicos na avaliação do N do solo e o método de extração mais difundido foi desenvolvido por Kjeldahl em 1883. Esse método fundamenta-se na conversão do N orgânico ($R-NH_2$) contido na amostra à NH_4^+ por meio de uma digestão sulfúrica, e a dosagem deste por meio da quantidade de NH_3 liberado pela destilação do digerido em meio alcalino. Assim o método processa-se em duas etapas: a digestão sulfúrica e a destilação (MENDONÇA & MATOS, 2017).

Na digestão, utiliza-se H_2SO_4 juntamente com uma mistura digestora composta por Na_2SO_4 , que tem por objetivo elevar o ponto de ebulição do H_2SO_4 , $CuSO_4$ e selênio metálico (limalha) em que o Cu e o Se atuam como catalisadores na oxidação da MO. A digestão é conduzida em chapa quente ou bloco digestor sob temperatura

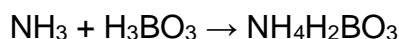
superior a 360°C, mas inferior a 410°C, pois a esta temperatura pode ocorrer alguma perda de NH₃. O tempo de digestão é de aproximadamente uma hora. Com a digestão sulfúrica o N orgânico (NH₂) é reduzido a NH₄⁺, formando (NH₄)₂SO₄.



A etapa de destilação deve ser conduzida em destilador com arraste de vapor. Para tanto, o produto da digestão Kjeldahl é tratado com uma solução fortemente alcalina, normalmente NaOH 40%, sob a injeção de vapor d'água, o que proporciona o desprendimento de NH₃, que é arrastado através do destilador.



A NH₃ é coletada em uma solução de H₃BO₃ com coloração vinho contendo os indicadores verde de bromocresol e vermelho de metila e é denominada de solução indicadora mista em ácido bórico. Ao receber o destilado, o pH da solução indicadora sofre um ligeiro aumento, o H₃BO₃ dissocia-se e reage com a NH₃ formando o borato de amônio.



Após esta reação, a solução assume a cor azul indicando a completa destilação da NH₃. A dissociação do ácido bórico é proporcional a quantidade de NH₃ coletada. Assim, fazendo-se a reconstituição do H₃BO₃ com a adição de H⁺, pode-se estimar quanto de NH₃ foi proveniente da destilação. Para tanto, processa-se uma titulação com uma solução ácida padrão, até que a solução indicadora em ácido bórico assuma coloração rósea.

2.3 ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO (NIR)

A região infravermelha (IV) foi descoberta ocasionalmente em 1800 pelo astrônomo W. Herschel, que ao estudar os efeitos térmicos nas diferentes regiões do espectro utilizando um termômetro, observou que a região espectral acima da cor vermelha fornecia grande potência calorífica. A técnica de análise por NIR, mesmo tendo um grande avanço nos dias atuais, ainda precisa de métodos analíticos tradicionais para se fazer a calibração do aparelho. Porém essa metodologia pode vir a substituir grande parte das análises convencionais em laboratório, garantindo assim a qualidade e especificidades necessárias para cada análise e por possuir inúmeras vantagens, tais como: agilidade no tempo de análise das amostras; menor custo;

abdicar do uso de reagentes e produtos químicos tóxicos ao meio ambiente e à saúde de quem está o manuseando; é uma análise múltipla de constituintes; e com menor necessidade de mão de obra para manuseá-lo (SANTOS, PEREIRA, & KORNDÖRFER, 2010).

A faixa espectral do NIR, está situada entre a região do visível e o IV médio, e vai de 13333.33 a 4000 cm^{-1} ou 750 a 2500 nm. É nesta região que a maior parte dos materiais orgânicos possuem propriedades de absorção da radiação, ou seja, um feixe de radiação é refletivo na região do NIR sobre a amostra, e assim, ligações covalentes, sobretons e bandas de combinação das ligações entre os componentes das substâncias orgânicas absorvem essa energia. Com isso, é calculado o quanto da emissão da luz que foi absorvida pela amostra com o objetivo de determinar os números e os tipos de ligações moleculares presentes na mesma (BELTRAME, 2014). O estudo de solos no NIR no Brasil é uma linha de pesquisa que vem crescendo desde 1980. A matéria orgânica, óxidos de ferro, granulometria, argilo-minerais, são os principais constituintes do solo que intervêm no comportamento espectral, mas o uso do NIR em solos é maior quando se trata de carbono orgânico (SANTOS, PEREIRA, & KORNDÖRFER, 2010; BELTRAME, 2014).

Para validação dos resultados obtidos no equipamento de leitura, utiliza-se a comparação dos resultados obtidos por métodos convencionais com o método proposto, o NIR, e a partir de realizações de testes, calibrações multivariadas, transformações, entre outros (SANTOS, PEREIRA, & KORNDÖRFER, 2010; SILVEIRA, 2009). No entanto, quanto mais complexa é a amostra por possuir um número elevado de interferentes presentes, mais difícil se torna a sua calibração, sendo necessária utilização de cálculos estatísticos mais robustos, que permitam a utilização de vários comprimentos de onda para determinação de uma única propriedade (COSTA, 2003). Com isso, é necessário e indispensável a utilização da quimiometria na aplicação do NIR.

2.3.1 Princípios da espectroscopia NIR

A espectroscopia estuda a interação da radiação eletromagnética com a matéria, e possui como um dos seus principais objetivos o estudo dos níveis de energia de átomos ou moléculas. Ao absorver radiação IV, as moléculas são excitadas

para atingir um estado maior de energia, e absorve então a radiação no IV que correspondem aos comportamentos vibracionais da molécula em estudo. Essa energia ao ser absorvida, modifica os movimentos vibracionais das ligações na molécula (BARBARA, 2013).

Para ocorrer a absorção da radiação infravermelha, é necessário que haja a variação do momento de dipolo elétrico da molécula como consequência de seu movimento vibracional, fazendo com que o campo elétrico alternante da radiação incidente interaja com a molécula, originando os espectros. Pode-se dizer então, que quando radiação eletromagnética incidente tem uma componente com frequência correspondente a uma transição entre dois níveis vibracionais é que origina o espectro de absorção no IV, sendo assim, o método é fundamentado na absorção seletiva da luz pelos compostos químicos presentes no material. Devido a constituição química da amostra, cada material (solo, madeira, carvão vegetal, alimento, etc) apresentará diferentes respostas ou absorção da luz. (COSTA, 2017; COSTA, 2003). Segundo COSTA (2003), os átomos em uma molécula nunca estão imóveis e a vibração dos átomos no interior de uma molécula apresenta energia coerente com a região do espectro eletromagnético correspondente ao IV.

O espectro IV de um composto químico é considerado uma de suas propriedades físico-químicas mais características e, por conta disto, a espectroscopia na região do IV tem extensa aplicação na identificação dos compostos.

2.3.2 Aplicações da espectroscopia NIR

No princípio do século XX já tinham sido publicados espectros de IV de centenas de compostos orgânicos e inorgânicos e haviam estudos e descobertas de algumas bandas de absorção sendo correlacionadas com a presença de determinados grupos funcionais nas moléculas. Porém somente por volta dos anos 1940 que começaram a aplicar a técnica por causa de suas limitações e dificuldades experimentais (ILHARCO, 1998).

A aplicabilidade e utilização da técnica NIR é bem ampla e pode ser usada como ferramenta analítica em determinações quantitativas e qualitativas de diversos materiais, sendo eles: indústrias de petróleo, têxtil, carvão, cosméticos, polímeros, química, tintas, farmacêuticas, alimentos e o estudo que vem se aprimorando cada

vez mais é no estudo de solos. Para utilização do IV em determinações qualitativas, deve-se saber que ele fornece: a informação estrutural da molécula em estudo, a absorção de cada tipo de ligação e cada tipo de ligação presente na molécula. Uma importante aplicação do IV, mas ainda bem menos utilizada, é para análise quantitativa de amostras, assim, o espectro IV pode servir para moléculas da mesma forma que impressões digitais servem para seres humanos.

Beltrame (2014) afirma que “deve-se salientar que a eficácia da espectroscopia de NIR é dependente de métodos analíticos tradicionais, visto que é necessária a sua calibração a partir dos resultados desses”. Porém sua aplicação está sendo recentemente bastante explorada devido a: agilidade nas análises, aplicação de métodos quimiométricos, baixo custo, método não invasivo, não destrutivo, pois não necessita de reagentes químicos, além de não precisar de pré-tratamento das amostras. Segue na tabela 1 alguns trabalhos com aplicação do NIR em análise de Solos.

Tabela 1 – Resumo dos trabalhos relacionados a técnica NIR e análise de solos

REFERÊNCIA	ANÁLISE	R ²	RMSEC	RMSEP
RMSE				
BELTRAME, 2014	COT	0,8710	1,9427	1,9795
DOTTO, et al. 2018	COT	0,54 – 0,82	7,7 – 4,8	
SANTOS, 2011	MO	0,80 – 0,84	8,1	10,3
SOUZA, 2014	COS	0,83	0,8 – 5,0	0,9 – 6,2
LOPEZ, 2009	Pedologia quantitativa	0,57 – 0,66	0,8 – 1,69	
SATO, 2013	COT	0,80 - 0,88	2,85	3,70
FELIX, et al. 2016	COT	0,86	2,09	
FELIX, et al. 2016	NT	0,97	0,10	
SEGNINI, et al. 2014	COT	0,99	16 %	
KUSUMO, SUKARTONO, BUSTAN, 2018	COT	0,76	0,151	
MARCHÃO, R. L., BECQUER, T., BRUNET, D, 2011	COT	0,91 calibração 0,69 validação	1,64	2,12
	NT	0,11 calibração 0,58 validação	0,11	0,23

RMSEC - erro médio quadrático calibração. RMSEP - erro médio quadrático da predição. RPIQ - relação entre o desempenho e a distância interquartil.

2.4 ANÁLISE MULTIVARIADA

A análise multivariada corresponde a um grande número de métodos e técnicas que utilizam, conjuntamente, todas as variáveis na análise teórica do conjunto de dados obtidos (COSTA, 2017).

Para realização da análise multivariada, é necessário a organização dos dados em um arranjo ordenado de linhas e colunas, constituindo uma matriz X ; a visualização e verificação da distribuição desses dados por meio de gráficos, o pré-tratamento de dados para reduzir as variações indesejáveis que não foram removidas durante a aquisição dos dados e que não serão eliminadas naturalmente durante a análise, mas que podem influenciar os resultados finais. Para o pré-tratamento consequentemente deve-se fazer a transformação e o pré-processamento dos dados (FERREIRA, 2015), que é realizado através da estatística multivariada. Existem vários e diferentes métodos de análise multivariada, que pode ser qualitativa, quando se pretende classificar ou explorar os dados ou quantitativa, quando se pretende estimar variáveis das amostras, e que são adotados de acordo com os objetivos da pesquisa (COSTA, 2017).

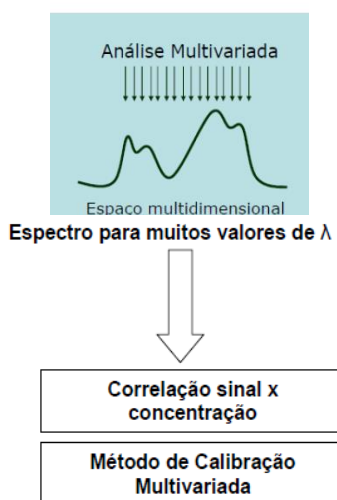


Figura 1- Ilustração do procedimento de calibração multivariada
Fonte: MONTEIRO, 2011

Com a expansão de ferramentas obtidas na química, que demandam utilização da estatística multivariada, nasceu uma nova disciplina designada por Svante Wold em 1972, de Quimiometria que é definida por TEÓFILO (2013, p.1) como a aplicação de quaisquer métodos matemáticos e estatísticos para o tratamento de dados

químicos [...]. FERREIRA (2015, p.25 apud CHRÉTIEN, 2003) fala que devido à interdisciplinaridade da quimiometria, uma nova definição foi proposta recentemente por chrétien: “quimiometria é a administração e processamento de informações de natureza química”. E conclui que a disciplina é uma interseção de três grandes áreas: química, matemática e a estatística. Atualmente a quimiometria está em bastante uso e seu estudo está sendo espalhado com agilidade em diversas áreas. Cada vez mais suas metodologias são aperfeiçoadas e direcionadas a resolver cada vez mais os problemas de origem química, principalmente na área da analítica. A química analítica foi a área mais beneficiada com a aplicação da quimiometria, pois o desenvolvimento da instrumentação analítica está interligado aos programas computacionais, e frequentemente surgem ou se modificam novos algoritmos, que conseqüentemente dependem e são empregados por ambientes computacionais (SOUZA, 2014; SOUZA e POPPI, 2012).

São vários os métodos quimiométricos existentes, e suas aplicações dependem do ambiente do problema que se deseja resolver, ou à informação que se pretende obter. Esses métodos podem ser classificados, de maneira geral como, reconhecimento de padrões, calibração multivariada, resolução multivariada de sinais e planejamento de experimentos. As etapas de análise são de extrema importância para que a interpretação e validação dos dados seja mais precisa (SOUZA, 2014).

Nessa dissertação será abordado brevemente alguns dos métodos quimiométricos aplicados neste trabalho, tais como a calibração multivariada, regressão por quadrados mínimos parciais, e validação cruzada.

2.4.1 Calibração Multivariada

O objetivo final da maioria das análises de dados multivariados é desenvolver modelos para quantificar uma propriedade de interesse (FERREIRA, 2015). O termo calibração para a quimiometria é interligado à construção de um modelo que relaciona resultados obtidos a partir de determinadas amostras às concentrações ou propriedades adquiridas através de métodos tradicionais destas amostras (TEÓFILO, 2013). Ou seja, “é um modelo que aborda o problema por intermédio da construção de modelos funcionais e algorítmicos sugeridos por uma grande quantidade de medidas experimentais” segundo (FERREIRA 2015, p.251). Seu caráter é empírico,

visto que é consequência de observações e medidas quantitativas extraídas do comportamento de todo o sistema, e têm como objetivo principal supor concentrações ou propriedades desconhecidas com base de respostas obtidas a partir de novas amostras.

A princípio, o modelo é calibrado com um conjunto definido de amostras de referência para as concentrações ou propriedades de interesse. As novas amostras, de onde serão obtidas as respostas previstas, normalmente apresentam as mesmas características das amostras de referência. Neste modelo, que vai possuir um conjunto de dados que constitui o conjunto de calibração, as respostas obtidas são variáveis que são designadas como preditoras, ou como variáveis independentes e o outro conjunto de dados contém os valores de algumas propriedades de interesse de cada uma das amostras de referência, ou seja, contém as concentrações ou propriedades são definidas como variáveis dependentes (FERREIRA, 2015; TEÓFILO, 2013).

Normalmente para realizar uma calibração multivariada, utiliza-se muitas variáveis, respostas instrumentais (um vetor de dados), para quantificar alguma propriedade de interesse (BELTRAME, 2014). Logo, os modelos construídos são fundamentados em um limitado conhecimento do investigador em relação ao princípio, mas onde uma grande quantidade de dados pode ser obtida, e são conhecidos como modelos flexíveis, pois, relacionam as respostas com as concentrações/propriedades conhecidas. Com a chegada e a vasta aplicabilidade da quimiometria, foram criados vários modelos que exigem certo tipo de arrumação das respostas para serem construídos. O arranjo utilizado nesse trabalho é o denominado de segunda ordem que utiliza muitos vetores de mesmo comprimento para cada amostra *i.e.*, uma matriz (TEÓFILO, 2013).

Para desenvolver modelos com calibração multivariada, necessita-se aplicação de duas etapas: a calibração e a validação. Na primeira etapa, o qual corresponde a calibração, as relações entre sinais analíticos são dispostos em uma matriz e as concentrações dos analitos são entendidos a partir de referências representativas para o conjunto de amostras a serem analisadas (SANTOS, 2011).

Para realizar calibração multivariada, conseqüentemente, trabalha-se com uma quantidade de dados extensa, e objetiva-se transpassar essas informações para outro sistema com menos variáveis através da ACP (do inglês “*principal component analysis*”). A proposta da ACP é expressar informações mais significativas contidas nas variáveis originais em um pequeno número de novas variáveis, as CP

(componentes principais). O número máximo de CP obtidos é igual ao número de vetores de dados utilizados, porém nem todas contém informações pertinentes à análise, ou seja, algumas CP's pode conter ruídos inerente aos dados. Quanto maior o número de CP, maior pode ser a presença do ruído proveniente do conjunto de calibração, levando assim a um aumento na variância da predição. Porém, ao adicionar poucos fatores no modelo, pode-se deixar de incluir informações relevantes levando a conter erros de previsão. Sendo assim, é necessário a determinação do número ideal de componentes principais que resulte na melhor previsão possível (SANTOS, 2011).

Dentre os modelos de regressão mais empregados em química analítica, com o intuito de construir modelos de calibração multivariada, o mais utilizado é a regressão por mínimos quadrados parciais (BELTRAME, 2014), descrito com maiores detalhes na seção 2.4.4.1.

2.4.1.1 Regressão por quadrados mínimos parciais (PLS)

De acordo com SOUZA (2014, p.11) “certamente a regressão por mínimos quadrados parciais [...] é um dos métodos mais proeminentemente empregados em Quimiometria, sendo considerado o método de calibração multivariada tido como referência para este tipo de aplicação”. FERREIRA (2015, p.346) também menciona que “é um método tendencioso que emprega análise de fatores e é de longe o método de regressão mais popular em quimiometria”. Já TEÓFILO (2013, p.45) diz que “de todos os métodos multivariados [...] o mínimos quadrados parciais é talvez o menos entendido. Entretanto, ele é amplamente usado [...]”.

O método dos mínimos quadrados parciais (do inglês “Partial Least Squares” - PLS) emprega princípios que capturam a maior variância dos dados, permitindo uma correlação máxima entre as variáveis originais e as previstas (SANTOS, 2011). São três as etapas de realização do método: na primeira monta-se as matrizes com as variáveis dependentes (matriz Y) e independentes (matriz X), a segunda elimina informações que não são proeminentes ao sistema, através de combinações lineares das colunas da matriz X feita pela análise das componentes principais, e a terceira e última etapa obtém-se a relação entre as matrizes X e a matriz Y que possui a propriedade de interesse (FIDÊNCIO, 2001). Trata-se de corrigir distorções

causadoras de variação na aquisição dos espectros que tendem a ocultar a informação útil dos dados (SANTOS, 2011). Conforme os seguimentos, possui o objetivo de aproximar o ambiente das medidas originais por um de tamanho reduzido, mas com a possibilidade de que ocorra alguma interferência na decomposição da matriz X direcionando-a para uma solução cujo alvo é a propriedade de interesse (FERREIRA, 2015).

Existem vários tipos de pré processamentos que produzem diferentes transformações nos dados, os que foram utilizados nesse trabalho foram: normalização e ponderação dos espectros, pois sempre vão conter ruídos, que são pequenas flutuações aleatórias e rápidas no espectro; suavização utilizando o algoritmo de Savitzky-Golay; inclinação da linha de base pois este influencia medidas como a altura e a área do pico (1^a e 2^a derivadas); correção multiplicativa de sinal (MSC - Multiplicative Scatter Correction) que é uma forma de modelar o sinal de um espectro como uma função linear de um espectro de referência ou o espectro médio (DEHANOV, 2011); e a variação normal padrão (SNV - Standard Normal Variate).

Pode-se calcular alguns parâmetros com os dados resultantes dos modelos de calibração, que irão indicar a eficiência dos mesmos ajudando assim critério de seleção dos melhores ajustes. São alguns aplicados nesse trabalho: coeficientes de correlação (R) ou determinação (R^2), número de fatores mínimos quadrados parciais, valores de erro padrão na calibração e da validação, raiz do erro padrão médio (RMSE), relação entre o desempenho e a distância interquartil (RPIQ). O valor de R^2 pode variar entre 0 e 1 e representa a correlação dos pontos da curva de calibração. Quanto mais próximo de 1, indica que os valores seguem um modelo bastante linear e a relação é bem explicada para o modelo proposto, caso contrário, deve-se avaliar outro modelo (COSTA, 2017).

O filtro Savitzky-Golay (SG) foi criado em 1964 e é um método de pré-tratamento bastante utilizado nos dias atuais, pois é um algoritmo de filtragem e suavização de dados, ou seja, tem como função liminar os ruídos como a linha de base-deriva, inclinação, reversão e assim por diante (CHEN et al, 2012; NISHIDA, 2017). O filtro contém diferentes modos de suavização, e emprega a capacidade de ajuste de regressão polinomial para melhorar os resultados de suavização dos espectros (DEHANOV, 2011), *i.e*, um polinômio de ordem baixa é ajustado aos pontos da janela e utilizado para recalcular o ponto central (SOUZA, 2014).

A finalidade da primeira derivada ($D1$), segundo DEHANOV (2011, p.22) é

“eliminar a linha de base, mantendo os detalhes espectrais necessários para correlacionar absorvâncias e concentrações”. A D1 reduz a sobreposição de um pico medindo o declive da tangente à curva do gráfico nos diferentes pontos, relacionando a variação da absorvância (dA) versus a variação do número de onda ($d\tilde{\nu}$). A obtenção do espectro da primeira derivada é colocando $dA/d\tilde{\nu}$ versus $\tilde{\nu}$, onde a anulação ocorre no ponto referente ao número de onda máximo do espectro de absorção de ordem zero (DONATO et al, 2010). Ao percorrer o gráfico do espectro, os sinais do valor da derivada variam (DEHANOV, 2011). É negativo onde a absorção decresce e positivo onde a absorção aumenta (DONATO et al, 2010). Ou seja, ao observar os valores subindo indo ao topo do pico, o valor da derivada vai coincidir com a inclinação do gráfico, ficando ambos positivos. No máximo onde pode ocorrer uma pequena região plana que fica ao topo do pico, a inclinação é zero e o valor da derivada também. Ao percorrer o gráfico em direção à base, *i.e.*, descendo o pico, a inclinação vai ser negativa e o valor da derivada também. Sendo assim, o valor da derivada acompanha a inclinação do gráfico do espectro. Ao mudar a inclinação faz-se com que a derivada de um pico tenha parte positiva, passe pelo zero no local do número de onda correspondente ao pico, e logo tenha uma parte negativa. Com esse intuito, as derivadas são usadas, por vezes, para localizar as posições de picos nos espectros (DEHANOV, 2011).

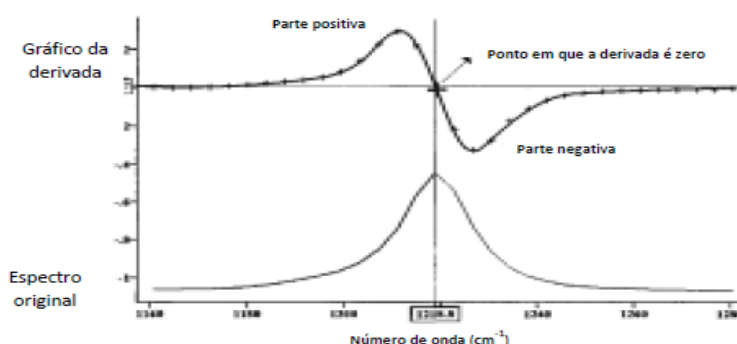


Figura 2 – Gráfico que ilustra o comportamento da primeira derivada
Fonte: (DEHANOV, 2011)

Ao gerar espectros, pode ocorrer que a linha de base diversifique um dado valor de espectro para espectro, fazendo com que o deslocamento da linha de base seja um problema ao medi-los. Essa variação nas linhas de bases vão produzir variações indesejadas nas medidas de absorvância e só é removido utilizando a primeira derivada. O intuito é fazer com que a linha de base passe a ser zero mas que as

informações das áreas dos picos após a aplicação da primeira derivada permaneça conservada, mantendo a correspondência quantitativa entre os espectros e a concentração em espectros derivados (DEHANOV, 2011).

A segunda derivada (D2) é formalmente conhecida como a derivada da primeira derivada e mede a profundidade ou quão curva é a função. Quando o pico está em posição virado para cima, a concavidade é positiva. Onde o pico é mais alto e consequentemente mais íngreme, a concavidade é zero e passa por um ponto de transição chamado ponto de inflexão. O topo de um pico vai possuir a concavidade negativa e voltada para baixo. Devido a isto,

[...] a segunda derivada de um pico tem uma parte positiva, passa por zero no ponto de inflexão, começa então uma parte negativa, passa por zero novamente no outro ponto de inflexão e volta a ter uma parte positiva novamente até voltar a ser zero, pois é esse o valor da linha de base. O valor mínimo da parte negativa da derivada do pico corresponde ao máximo do pico. É devido a este comportamento que a segunda derivada é usada para “separar” grupos de picos sobrepostos (DEHANOV, 2011, p.23)

O intuito da D2 é similar ao de primeira que é eliminar as linhas de base e junto a isto eliminar a inclinação do espectro do conjunto de dados espectrais. As informações da área dos picos após a aplicação da segunda derivada também é mantida preservando a correlação entre a medida espectral e os valores de concentração (DEHANOV, 2011).

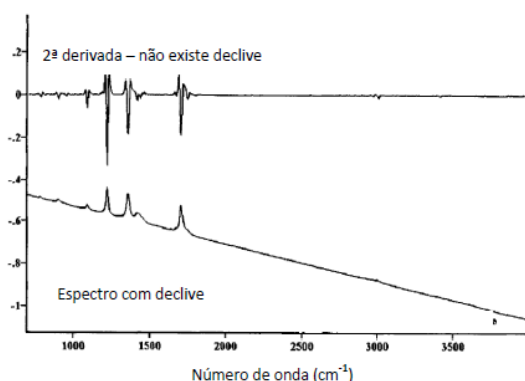


Figura 3 - Gráfico de um espectro contendo declive e gráfico de um espectro de segunda derivada onde o declive foi eliminado

Fonte: (DEHANOV, 2011, p.23).

Ainda na etapa de calibração dos modelos, ocorre a eliminação de amostras anômalas, que são as amostras com comportamento diferente das demais no conjunto de amostras, denominadas *outliers*. Essa etapa é essencial para o êxito da análise multivariada, pois os dados obtidos nas análises experimentais pelos métodos

tradicionais, são muitas vezes confusos, incompletos, ou até mesmo errôneos. Para obter uma melhoria e correção dessas falhas nos dados, é preciso como primeiro passo, realizar um pré-processamento para assegurar que as amostras formem um conjunto homogêneo. *i.e*, essa técnica possui como principal objetivo eliminar informações indesejáveis e tornar a matriz de dados mais coerente e mais uniforme possibilitando a subsequente análise exploratória do conjunto de dados com eficiência (DEHANOV, 2011; FERREIRA, 2015; SOUZA, 2014, SANTOS, 2011). Os *outliers* podem ocorrer por diversas razões. Pode-se dizer erros de digitação, erros de execução da análise, erros instrumentais, dentre eles, quando o espectro da amostra apresentar baixa qualidade tornando-o não interpretável. O ideal é o analista alertar-se e repetir a amostra, ou caso contrário, excluí-la do banco de dados. (FERREIRA, 2015; TEÓFILO, 2013).

Para realizar o pré - processamento nos dados, é necessário executar vários passos e conhecer o que se vai aplicar, pois existem vários esquemas possíveis para realização do mesmo. Sendo assim, deve-se inspecionar sempre os dados antes e após o pré-processamento para avaliar se a informação relevante foi mantida enquanto foi removido.

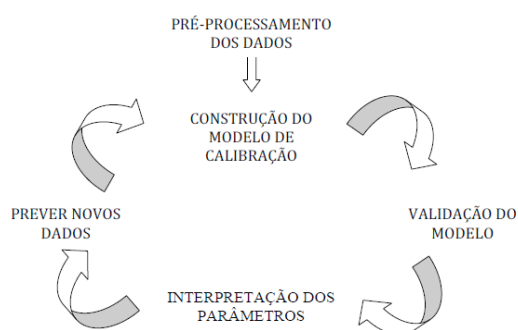


Figura 4 - Processo geral de construção de modelos de regressão
Fonte: (MONTEIRO, 2011)

Após a calibração, é feita a validação dos modelos que consiste no melhoramento da relação que descreve espécies de interesse na avaliação da habilidade preditiva do modelo construído (SANTOS, 2011), descrito com maiores detalhes na seção 2.3.4.2.

2.4.1.2 Validação do modelo

A validação da calibração multivariada pode ser realizada em duas maneiras: validação externa e a validação interna ou cruzada (do inglês *cross validation*). O modelo aplicado na dissertação foi a validação cruzada, que é feita durante a sua construção. *i.e*, todos os padrões preparados são empregados no cálculo do modelo e são utilizados também como amostras para o teste do modelo construído. Algumas amostras são separadas, um modelo é construído com as amostras restantes, e a hipótese é feita com as amostras separadas anteriormente. Durante esse processo é possível avaliar melhor a quantidade de constituintes do modelo, detectar amostras desiguais às outras do banco de dados e apurar com antecedência sua robustez e capacidade de previsão, *i.e*, faz-se a seleção das amostras do conjunto de calibração. A partir desse modelo construído, ocorre a validação para amostras externas que se conhece a propriedade de interesse y e os respectivos dados em X , mas que ainda não foram utilizadas para a etapa de calibração. O conjunto de amostras é denominado conjunto de teste ou conjunto externo de validação e possui como único objetivo testar a capacidade preditiva do modelo (TEÓFILO, 2013; FERREIRA, 2015; SANTOS, 2011).

3 MATERIAL E MÉTODOS

Foram obtidas 901 amostras de solos referentes a propriedades rurais atendidas pelo Laboratório de Solos do CCAE-UFES. Essas amostras constituíram o banco de dados para análise no NIR e dentre estas, 621 foram utilizadas para análise de COT e 280 para análise de NT. Essas amostras são originárias de diversas cidades do Sul do estado, como: Cachoeiro de Itapemirim, Alegre, Guaçuí, Jerônimo Monteiro, Muqui, Mimoso do Sul, Caparaó, Castelo, Alfredo Chaves, Ibitirama, Iúna, Vargem Alta, Atílio Vivácqua, São José do Calçado, Muniz Freire, Presidente Kennedy, entre outras. Todas as amostras utilizadas foram secas ao ar ou em estufa 40°C (Terra Fina Seca ao Ar - TFSA), trituradas e passadas em peneira de 0,2 mm (60 mesh).

3.1 CARBONO ORGÂNICO TOTAL DO SOLO

Para obtenção do teor de COT, utilizou-se o método Walkey-Black modificado via úmida, com auxílio de bloco digestor como aquecimento externo (MENDONÇA & MATOS, 2017).

Pesou-se aproximadamente 0,1; 0,3 e 0,5 g para diferentes tipos de solos, ou seja, de acordo com os teores de CO presente no solo foi necessário quantidades também diferentes para realização da análise. O material foi transferido para tubos de digestão. Adicionou-se 5 mL da solução $K_2Cr_2O_7$ 0,167 mol L⁻¹ com ajuda de uma pipeta volumétrica, em seguida foi acrescentado 7,5 mL de H₂SO₄ concentrado. Após o pré-aquecimento do bloco digestor até a temperatura de 170°C, os tubos foram colocados no bloco digestor e mantidos nessa temperatura por 30 minutos. Em seguida, deixou-se esfriar. O conteúdo de cada tubo foi transferido quantitativamente para erlenmeyers de 250 mL, utilizando-se água destilada suficiente para um volume final de cerca de 80 mL. Foram adicionados 0,3 mL da solução indicadora e titulou-se com a solução de sulfato ferroso amoniacal 0,2 mol L⁻¹ (Sal de Mohr).

Nas mesmas condições foram feitos pelo menos 6 brancos controles (5,0 mL de $K_2Cr_2O_7$ 0,167 mol L⁻¹ + 7,5 mL de H₂SO₄ concentrado). Três destes brancos foram levados para a digestão, e três permaneceram sem aquecimento, em temperatura

ambiente. Este branco controle não aquecido foi importante para o cálculo do total de dicromato perdido no aquecimento na ausência da amostra.

A porcentagem de C foi calculada com base no volume da solução de Sal de Mohr gasto na titulação da amostra (V), do branco aquecido (Vba) e do branco não aquecido (Vbn), conforme as equações:

$$A = [(Vba - Vam) (Vbn - Vba) / Vbn] + (Vba - Vam) \quad (1)$$

em que:

Vba = volume gasto na titulação do branco controle com aquecimento;

Vbn = volume gasto na titulação do branco controle sem aquecimento;

Vam = Volume gasto na titulação da amostra.

$$CO \text{ (g/Kg)} = \frac{(A) \text{ (molaridade Sulf. Ferroso)}(3)(1000)}{\text{Peso da amostra (mg)}} \quad (2)$$

em que:

3 = resultado da relação entre o número de mols de $Cr_2O_7^{2-}$ que reagem com Fe^{2+} (1/6), multiplicado pelo número de mols de $Cr_2O_7^{2-}$ que reagem com o C (3/2), multiplicado pela massa atômica do C (12).

1000 = fator de conversão de unidade (mg mg^{-1} para g Kg^{-1}).

3.2 NITROGÊNIO TOTAL DO SOLO

O NT do solo foi determinado pelo método Kjeldahl, que se fundamenta na conversão do N orgânico ($R-NH_2$), pois cerca de 95% do N total do solo está na forma orgânica com a oxidação da MO, a primeira forma de N liberada é NH_4^+ por meio de digestão sulfúrica, e a dosagem deste por meio da quantidade de NH_3 liberado pela destilação do digerido em meio alcalino (MENDONÇA & MATOS, 2017). Assim o método processa-se em duas etapas: a digestão sulfúrica e a destilação. Para digestão sulfúrica utilizou-se em média 0,2 a 0,5 g de solo, podendo ser mais ou menos a quantidade de acordo com o tipo do solo, e colocado em tubo de digestão de 100 mL. Adicionou-se 1 mL de H_2O_2 30%, 2,0 mL de H_2SO_4 concentrado e deixou

esfriar por 15 minutos. Após, adicionar 0,7 g de mistura de digestão composta por Na_2SO_4 , CuSO_4 e selênio metálico, colocou as amostras no bloco digestor, elevou-se a temperatura até 250°C manteve-a por 20 minutos. Em seguida, se elevou a temperatura a 350°C , mantendo-a por 2 horas. Deixou-se esfriar. Para o processo de destilação, adicionou-se 5 mL de água deionizada, agitou-se em agitador de tubos, conectou-se o tubo de digestão ao destilador Kjeldahl, e adicionou-se 10 mL de NaOH 10 mol/L sob a injeção de vapor d'água, o que proporciona desprendimento de NH_3 , que é arrastada através do destilador. Destilou-se o NH_3 em 5mL de ácido bórico contendo os indicadores verde de bromocresol e vermelho de metila. Ao receber o destilado o pH da solução indicadora sofre um ligeiro aumento, o H_3BO_3 dissocia-se e reage com o NH_3 formando o borato de amônio. Para tanto, processou-se uma titulação com uma solução ácida padrão HCl 0,05 mol/L até que a solução indicadora de ácido bórico assumisse uma coloração rósea (MENDONÇA & MATOS, 2017).

Feito esse procedimento, foi feito os seguintes cálculos para determinação do parâmetro desejado:

$$N \text{ (g Kg}^{-1}\text{)} = \frac{(V_{\text{am}} - V_{\text{b}}) \times [\text{H}^+] \times 14}{\text{Peso do solo (g)}} \quad (3)$$

Em que:

V_{am} = Volume de HCl gastos na titulação da amostra

V_{b} = Volume de HCl gastos na titulação do branco

$[\text{H}^+]$ = concentração real do ácido clorídrico (mol L^{-1})

14 = massa molecular do N (g mol^{-1})

3.3 ESPECTROSCOPIA DE INFRAVERMELHO PRÓXIMO - EQUIPAMENTO NIR

Os espectros NIR de reflectância das amostras foram adquiridos em um espectrômetro Interline Quant QIA 1020 TYPE C modelo MB3600 com transformada de Fourier.

Uma amostra em branco foi utilizada para estabelecer a linha de base. O intervalo espectral gerado encontra-se no comprimento de onda compreendido entre 645 e 5024 nm, que corresponde ao número de onda de 15503.88 cm^{-1} e 1990.45 cm^{-1} . O que tornou possível determinar com precisão os componentes, exibindo os espectros, as informações e tendências das amostras imediatamente. Para

construção e validação dos modelos de predição, utilizou-se os comprimentos de onda na faixa do NIR, entre 1.000 e 2.500 nm, ou seja, 10000.00 cm^{-1} e 4000 cm^{-1} com uma resolução de 16 cm^{-1} . A faixa de número de onda fora desse intervalo foi descartada por apresentar alta relação sinal/ruído. As intensidade de cada passo da leitura obtidas em refletância obtido para cada intervalo de número de onda de 7.148 cm^{-1} , armazenadas e processadas em arquivos disponíveis no software FemWin900.

3.4 AVALIAÇÃO DE OUTLIERS

A análise de componentes principais (ACP) foi realizada em cada amostra de cada conjunto de dados e a partir dos escores foi realizada uma análise de agrupamento baseado na distância euclidiana e grupos com distância de Mahalanobis (H) maior que 3 foram considerados outliers e removidos do conjunto de dados permitindo gerar a homogeneidade dos dados e excluir dados com problemas na aquisição dos dados espectrais.

3.5 SELEÇÃO CONJUNTO DE CALIBRAÇÃO E VALIDAÇÃO

3.5.1 Carbono Orgânico Total

Após a retirada dos outliers, setenta por cento de todo o conjunto de dados ($n = 584$) foi escolhido por amostragem aleatória e usado para o conjunto de calibração ($n = 410$). Os restantes 30% foram usados para o conjunto de validação ($n = 174$). Amostras independentes de solo foram usadas em cada conjunto. A separação dos grupos foi realizada pelo método K-média com o objetivo de gerar grupos com maior semelhança das análises estatística descritivas.

3.5.2 Nitrogênio Total

Setenta por cento de todo o conjunto de dados ($n = 211$) foi escolhido por amostragem aleatória e usado para o conjunto de calibração ($n = 148$). Os restantes

30% foram usados para o conjunto de validação ($n = 63$). Amostras independentes de solo foram usadas em cada conjunto. A separação dos grupos foi realizada pelo método K-média com o objetivo de gerar grupos com maior semelhança das análises estatística descritivas.

3.6 PRÉ-TRATAMENTOS DOS ESPECTROS

Foram realizados neste trabalho alguns métodos de pré-processamentos normalmente empregados em espectroscopia NIR: Primeira e segunda derivada; remoção contínua; variação normal padrão (SNV); Savitzky-Golay; Savitzky-Golay com aplicação da primeira derivada; Correção multiplicativa de sinal (MSC). Algumas figuras de mérito foram empregadas conjuntamente para avaliar o desempenho dos modelos de calibração multivariada: Raiz quadrada do erro médio, R^2 , relação entre o desempenho e a distância interquartil, e linearidade de resíduos (bias).

3.6.1 Ajuste do modelo de mínimos quadrado parciais

O modelo de regressão aplicado foi o regressão por mínimos quadrados parciais para construir os modelos de calibração a partir das amostras de calibração, e avaliar a desempenho dos métodos de pré-processamento. Toda a modelagem foi realizada com a linguagem de programação R pacote mínimos quadrados parciais (R Core Team, 2016) e parâmetros Jackknifing com ajuda e complemento do programa Statistica (STATSOFT, 2008) e Sigma Plot (SYSTAT SOFTWARE Inc, 2006). Este procedimento foi realizado com a normalização de variâncias (SNV) e correção multiplicativa de sinal (MSC), que buscaram reduzir a variação na inclinação de um espectro causada pelo efeito de espalhamento e por variações nos sinais das curvas, removendo a tendência curvilínea de um espectro.

Posteriormente, foi realizada a validação externa, baseada em amostras que não participaram da calibração. Três medidas estatísticas foram utilizadas nos métodos multivariados para avaliar o modelo ajustado: o coeficiente de determinação (R^2) (Eq. 4) com $\beta_0=0$ e $\beta_1=1$, erro quadrático médio (RMSE) (Eq. 5), e razão entre o desempenho para o intervalo interquartil (RPIQ) (Eq. 6). R^2 fornece a porcentagem da

variância explicada pelo modelo e é a medida mais largamente usada e relatada da avaliação do ajuste do modelo. O erro quadrático médio é comumente usado para medir a diferença entre o previsto e valores observados do modelo ajustado e é uma interpretação fácil estatística porque tem as mesmas unidades de dados. Essas diferenças individuais também são chamadas de residuais, e o erro quadrático médio da raiz serve para agregá-las em uma única medida de poder preditivo. Erro médio da raiz quadrada mede quantos erros há entre dois conjuntos de dados. Em outras palavras, erro quadrático médio compara um valor previsto e um valor observado ou conhecido. O intervalo interquartil é baseado em quartis, que melhor representam a disseminação da população. Para melhor explicação do intervalo interquartil para distribuição do conjunto de dados usando a distância interquartil, os conjuntos de amostras de solo devem possuir uma distribuição que não seja a distribuição normal.

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (4)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (5)$$

$$RPIQ = \frac{(Q3 - Q1)}{RMSE} \quad (6)$$

Teste para *bias*

$$SDV = \sqrt{\frac{\sum_{i=1}^{n_v} (\hat{y}_i - \bar{y})^2}{n_v}} \quad (7)$$

$$t_{bias} = \frac{|bias| \sqrt{n_v}}{SDV}$$

onde \hat{y} é os valores previstos, \bar{y} é a média do valor observado, y é o valor observado, n é o número de amostras com i igual a 1, 2, ... n , $Q1$ é a diferença entre o terceiro e o primeiro quartil ($Q3 - Q1$), $Q1$ é o valor em 25% das amostras, e $Q3$ é o valor em 75% das amostras.

Outra figura de mérito figura de mérito que pode ser empregada na avaliação dos modelos de calibração é o teste para erros sistemáticos (*t-bias*) (Eq. 7). Estes erros influenciam a exatidão das medidas, e são calculados pela diferença entre a média da população e o valor verdadeiro, sendo associados a todas as componentes

do erro que não são aleatórias, como exemplo, as configurações instrumentais (SOUZA, 2014).

Ao final do processo comparou-se os resultados dos dois métodos, NIR versus métodos de rotina, para caracterização, eficiência e a confiabilidade das análises de carbono orgânico total e nitrogênio total.

4 RESULTADOS E DISCUSSÃO

4.1 ESTATÍSTICA DESCRITIVA

4.1.1 Carbono Orgânico Total

Considerando os dados do conjunto total da amostragem, do conjunto de calibração e validação, o total dos valores de COT tiveram uma variação entre 4,30 e 52,80 (g Kg⁻¹) (Tabela 2). Os parâmetros estatísticos (média, variância, desvio padrão) desses grupos avaliados, foram semelhantes após a seleção grupos de calibração e validação pelo método K-média. A localização de uma área ativada através de uma análise estatística pode ser confirmada algoritmo de agrupamento K-média, ou seja, é um algoritmo muito utilizado para agrupamento de dados, fornecendo uma classificação de acordo com os próprios dados em análise, obtendo uma classificação por similaridade de grupos (clusters).

Tabela 2 – Estatísticas descritiva dos teores de COT (g Kg⁻¹) avaliadas em laboratório de rotina para o conjunto de calibração e validação

	AMOSTRAGEM TOTAL	CALIBRAÇÃO	VALIDAÇÃO
Média	15,73	15,48	16,34
Mínimo	4,30	4,30	6,30
Máximo	52,80	49,70	52,80
CV (%)	42,20	41,00	44,56
DP	6,64	6,35	7,28
Curtose	5,24	5,16	5,05
K-S	0,13	0,15	0,18

CV – Coeficiente de Variação; DP – desvio padrão; K-S - Kolmogorov-Smirnov

Mostra-se na Figura 5 (A1) e (A2) que a média dos teores para os dois conjuntos foram semelhantes, na Figura 5 (B1), o gráfico do conjunto de calibração tem tendência leptocúrtica, *i.e*, possui a curva da função de distribuição mais afunilada com um pico com mais intensidade do que a distribuição normal, assim como o conjunto de validação. Os valores $b_2 > 0$ (curtose) informados na Tabela 2 confirmam a tendência do achatamento da curva da função de distribuição. Com isso, percebe-se que os dados não seguem uma distribuição normal, e isso faz com que o resultado

da predição não seja tão eficiente.

Os conjuntos de calibração e validação usaram amostras independentes. No entanto, a semelhança entre os conjuntos revelou que a divisão aleatória dos grupos foram estatisticamente semelhantes e que a análise multivariada é adequada.

Os gráficos contidos na Figura 5 demonstram a estatística descritiva do COT para os conjuntos de calibração e de validação.

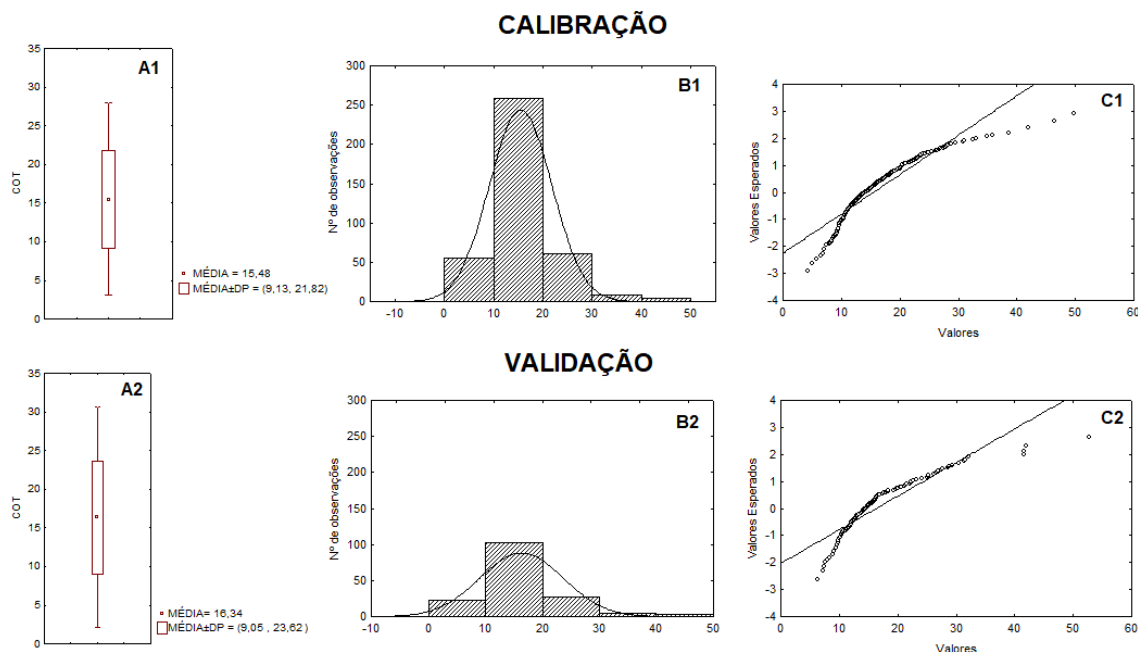


Figura 5 – Estatística descritiva dos conjuntos de calibração e validação para COT. (A1) e (A2) densidade populacional definida pela média e desvio padrão. (B1) e (B2) limite normal esperado para o número de observação pelo gráfico de curtose. (C1) e (C2) gráfico de dispersão.

A variabilidade dos teores desses atributos com CV, calculado pela razão entre o desvio padrão e a média, de 41,00 e 44,56% para o conjunto de calibração e validação respectivamente, reflete a amplitude de variação existente nos teores de COT para a região sul do Estado do Espírito Santo, *i.e*, a classificação dos solos da localidade vai de solos minerais que correspondem aos solos com teores <5% de COT, a solos >8% de COT. As Figuras 5 (C1) e (C2), nos dá a maneira de descrever como as probabilidades são associadas aos valores ou aos intervalos de valores de uma variável aleatória. Caso uma amostra apresente um valor menor que o mínimo ou maior que o máximo, a estimativa por NIRS nesta amostra para esta propriedade não será satisfatoriamente precisa.

4.1.2 Nitrogênio Total

Considerando os dados do conjunto total da amostragem, do conjunto de calibração e validação, o total dos valores de NT foram colocados entre 0,02 e 2,90 (g Kg⁻¹) (Tabela 3). Os parâmetros estatísticos (média, variância, desvio padrão) desses grupos avaliados, foram semelhantes após a seleção grupos de calibração e validação pelo método K-média.

Tabela 3 – Estatísticas descritiva dos teores de NT (g Kg⁻¹) avaliadas em laboratório de rotina para o conjunto de calibração e validação

	AMOSTRAGEM TOTAL	CALIBRAÇÃO	VALIDAÇÃO
Média	1,26	1,27	1,23
Mínimo	0,02	0,02	0,48
Máximo	2,90	2,90	2,34
CV (%)	34,70	35,72	32,16
DP	0,44	0,45	0,40
Curtose	1,00	1,23	-0,22
K-S	0,83	0,91	0,11

CV – Coeficiente de Variação; DP – desvio padrão; K-S - Kolmogorov-Smirnov

O valor de desvio padrão mostrou tendência dos dados mais homogêneos, pois seus valores estavam bem próximos de zero. Sua variância obteve valor elevado, pois indica o quanto os valores se desviaram da média. A média dos teores para os dois conjuntos foram semelhantes, e percebe-se na Figura 6 (B1), que o gráfico do conjunto de calibração tem tendência leptocúrtica, *i.e*, possui a curva da função de distribuição mais afunilada com um pico com mais intensidade do que a distribuição normal, em contrapartida, o conjunto de validação apresentou tendência platicúrtica, no qual a função de distribuição é mais achatada que a distribuição normal. Confirmam-se essas tendências do achatamento da curva da função de distribuição com os valores informados na Tabela 1, de $b_2 > 0$ (curtose), para o conjunto calibração e $b_2 < 0$ para o conjunto validação. Com isso, percebe-se que os dados não seguem uma distribuição normal, e isso faz com que o resultado da predição não seja eficiente.

Os conjuntos de calibração e validação usaram amostras independentes. No entanto, a semelhança entre os conjuntos revelou que a divisão aleatória grupos foram estatisticamente semelhantes e que a análise multivariada é adequada.

Os gráficos contidos na Figura 6 demonstram a estatística descritiva do NT para

os conjuntos de calibração e de validação.

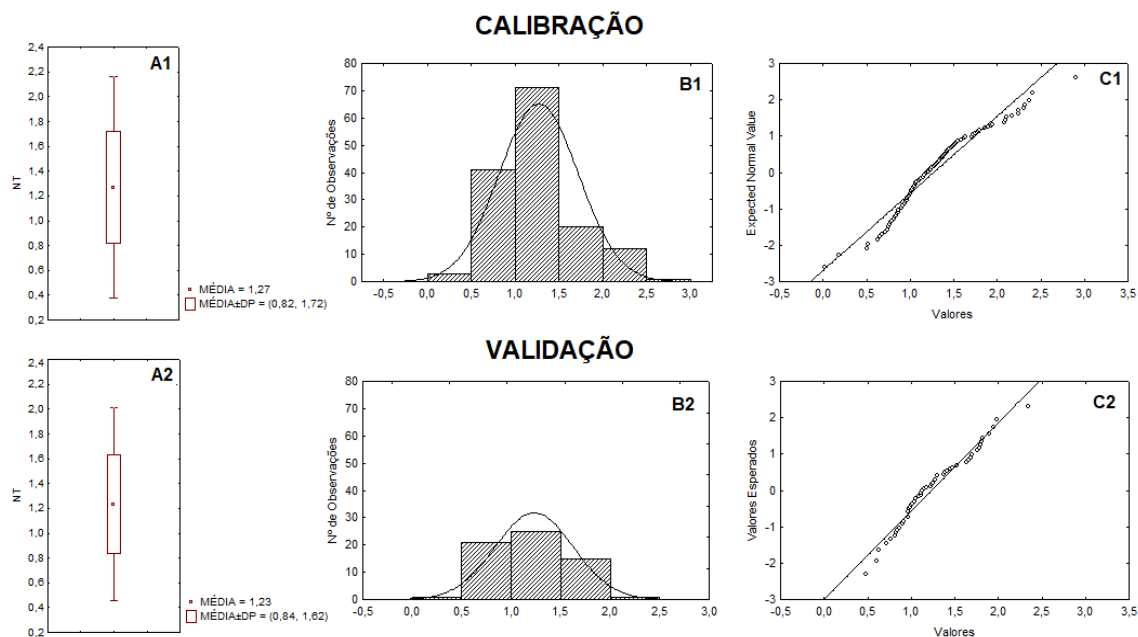


Figura 6 – Estatística descritiva dos conjuntos de calibração e validação para NT. (A1) e (A2) densidade populacional definido pela média e desvio padrão. (B1) e (B2) limite normal esperado para o número de observação pelo gráfico de curtose. (C1) e (C2) gráfico de dispersão.

A variabilidade dos teores desses atributos com CV, calculado pela razão entre o desvio padrão e a média, de 35,72 e 32,16% para conjunto de calibração e validação respectivamente, reflete a amplitude de variação existente em relação à média, nos teores de NT para a região sul do Estado do Espírito Santo. As Figura 6 (C1) e (C2) nos dá a maneira de descrever como as probabilidades são associadas aos valores ou aos intervalos de valores de uma variável aleatória. Caso uma amostra apresente um valor menor que o mínimo ou maior que o máximo, a estimativa por NIRS nesta amostra para esta propriedade não será satisfatoriamente precisa.

4.2 ESPECTROS NO NIR E CARACTERÍSTICAS DAS CURVAS DE REFLECTÂNCIA ESPECTRAL DO SOLO

Espectros das amostras foram coletados na faixa de 10.000 cm^{-1} e 4.000 cm^{-1} (espectro total), na figura 7 mostra-se os espectros médios coletados de todas as amostras.

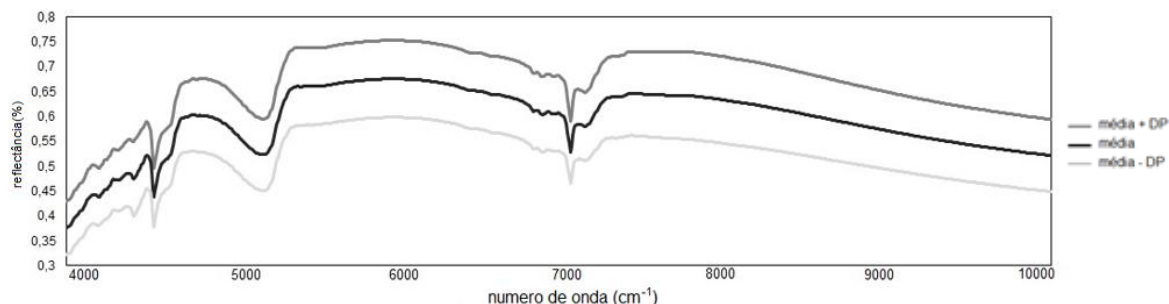


Figura 7 – Reflectância dos espectros totais na região NIR (4000 cm^{-1} – 10000 cm^{-1}) das amostras utilizadas para calibrar e validar os modelos

O padrão espectral do conjunto de dados (Figura 7) indica que além da presença de *outliers*, observa-se a existência de deslocamentos verticais de um espectro em relação ao outro nas matrizes de calibração e validação. Este efeito é indesejável do ponto de vista químico, pois carrega informações que não estão relacionadas com o sinal analítico o qual se está calibrando (SOUZA, 2014). Estas diferenças espectrais podem existir por diversas maneiras, tais como, tamanho das partículas de solo presentes no amostrador, concentração de argila e NT, erro amostral ou erro instrumental, entre outros. Deste modo, a remoção dos *outliers* foi caracterizada não apenas pela retirada de amostras com características indesejáveis instrumentalmente, mas também por remover amostras com baixa representatividade nos conjuntos de calibração e validação.

A seleção das amostras para compor os conjuntos de calibração e validação baseou-se na representatividade espectral e suas distâncias utilizando o algoritmo K-média, para que os dois conjuntos contemplassem a variabilidade dos solos em estudo mas com similaridade entre eles. O algoritmo de forma iterativa atribui os pontos de dados ao grupo que representa a menor distância, ou seja, ao grupo de dados que seja mais similar. Deste modo, é aplicado para realizar a seleção das amostras que irão compor o conjunto de calibração, uma vez que procede a seleção das amostras de maior variabilidade, *i.e.*, as amostras mais aleatórias do conjunto total. Assim, cada conjunto de dados foi inicialmente composto por 410 amostras para o conjunto de calibração e 174 amostras para o conjunto de validação para o teor de COT, e 148 amostras para o conjunto de calibração e 63 amostras para o conjunto de validação para o teor de NT, para ambos os métodos avaliados.

Uma importante característica para os modelos de calibração propostos serem satisfatórios, é a dimensionalidade dos dados das amostras e sua correlação. As

Figuras 8 e 9 relacionam o teor mínimo (4,30 g Kg⁻¹ COT e 0,02 g Kg⁻¹ NT) e máximo (52,80 g Kg⁻¹ COT e 2,90 g Kg⁻¹ NT), mostrados nas Tabelas 2 e 3, respectivamente, com os seus respectivos espectros, e confirma a grande variação das informações obtidas para o teor de COT e uma maior homogeneidade nos dados para o teor de NT. Percebe-se que na amostra com teor máximo de C, o espectro possui picos com menos intensidade de absorção sendo estes com bandas mais largas. O contrário ocorre na amostra com teor mínimo de C, que apresenta bandas mais justas e com picos mais acentuados. Isto ocorre devido aos tipos de estiramentos e vibrações dos grupos orgânicos contidos nestas amostras. Para o espectro de teores de NT, há uma maior semelhança nos picos de absorção, mantendo as bandas mais próximas e até sobrepostas, podendo ocorrer este fator visto que os teores analisados de NT não são tão distintos.

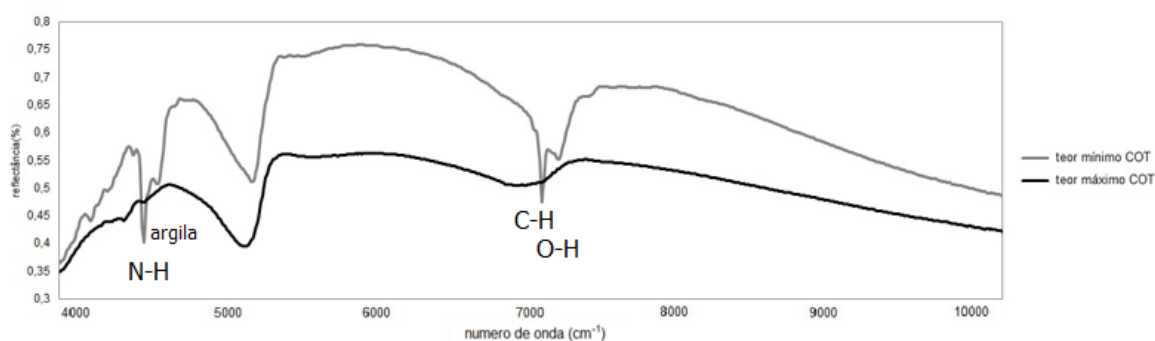


Figura 8 – Reflectância dos espectros na região NIR (4000 cm⁻¹ – 10000cm⁻¹) das amostras com teor mínimo e máximo de COT

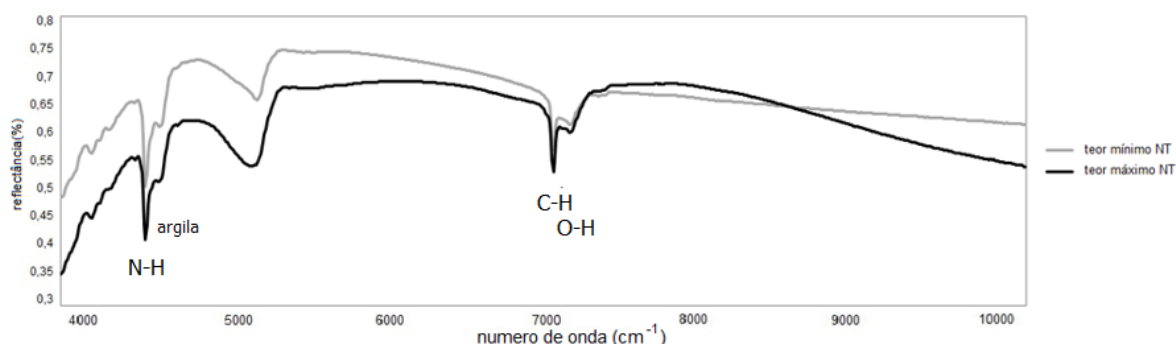


Figura 9 – Reflectância dos espectros na região NIR (4000 cm⁻¹ – 10000cm⁻¹) das amostras com teor mínimo e máximo de NT

Uma curva de reflectância espectral de cada amostra pode caracterizar a variabilidade das propriedades do solo. As amostras indicaram presença de curvas de reflectância distintas que foram associadas com diferentes formas e bandas de

absorção. Esta distinção pode ter sido principalmente causada pelo teor de carbono orgânico total e consequentemente os teores de argila e Nitrogênio contidos nas amostras.

Na comparação do conjunto de espectros é possível observar as três regiões, com distintas características de menor reflexão. Primeira região de $\sim 4500\text{ cm}^{-1}$ ($\sim 2222\text{ nm}$), a segunda na proximidade de 5200 cm^{-1} ($\sim 1920\text{ nm}$) e a terceira com banda na região de 7000 cm^{-1} ($\sim 1420\text{ nm}$).

Espectros de uma determinada amostra de solo é o resultado de energia por moléculas orgânicas que contém grande número de ligações do tipo C-H, N-H e O-H. Algumas características físicas como fator umidade, pode influenciar nos picos dos comprimentos de onda próximos a 7000 cm^{-1} , que ainda pode representar o primeiro sobreton de estiramento das ligações O-H, que estão relacionadas às moléculas de água ou de algum metal hidratado (SANTOS A. P., 2011; FELIX et al, 2016; SOUZA, 2014). O mesmo comportamento acrescentado de flexões das ligações pode representar o número de onda próximo a 5200 cm^{-1} . Ainda na região próximo a 7000 cm^{-1} , também indica diferenças de teores de caulinita das amostras, um dos principais minerais da fração argila dos solos, ou seja, esses picos podem ocorrer em razão do primeiro sobreton de moléculas O-H ligadas à caulinita. Sobretons de ligação C-H de grupos aromáticos, estão presentes na região próximo a 7000 cm^{-1} , assim como várias combinações de C-H a grupos alifáticos (4500 cm^{-1}) (SANTOS A. P., 2011; FELIX et al, 2016; SOUZA, 2014). Região de 4700 a 4500 cm^{-1} , sofre influência de grupos N-H aminas e amidas e também corresponde às ligações com C orgânico, que pode estar relacionada à presença de caulinita na amostra. O que quer dizer, que nesta região, pode ter a interferência do teor de argila no solo com teor de C orgânico (SANTOS A. P., 2011; FELIX et al, 2016; SOUZA, 2014).

No NIR para análise de solos, há uma dificuldade em observar uma relação direta entre número de onda e reflectância dos principais constituintes que influenciam seu comportamento espectral (matéria orgânica, argilo-minerais, granulometria e umidade), pois estes possuem características bem próximas um do outro (OLIVEIRA, 2015; SANTOS, 2011).

Os diferentes picos de absorção nos diferentes comprimentos de onda apresentados, podem estar ligados às diferentes combinações de espalhamentos de ligações orgânicas, às frações de argila que estão diretamente relacionados com os teores de COT do solo, e aos diferentes teores de NT que também tem interferência

nos teores de C.

4.3 PRÉ PROCESSAMENTO DOS DADOS

Foram realizados neste trabalho alguns métodos de pré-processamentos normalmente empregados em espectroscopia NIR: Primeira e segunda derivada; remoção contínua; variação normal padrão (SNV); Savitzky-Golay; Savitzky-Golay com aplicação da primeira derivada; Correção multiplicativa de sinal (MSC). Algumas figuras de mérito foram empregadas conjuntamente para avaliar o desempenho dos modelos de calibração multivariada: Raiz quadrada do erro médio, R^2 , relação entre o desempenho e a distância interquartil, e linearidade de resíduos (bias).

Conforme esperado, os métodos de pré-processamento aplicados minimizaram a influência de efeitos físicos nos espectros, os quais não trazem informações químicas sobre as amostras e introduzem variações espectrais, como o deslocamento da linha de base.

A Figura 10 mostra a média dos espectros brutos sem transformação e a Figura 11 (A), (B), (C), (D), (E), e (F) mostra os espectros pré processados por primeira derivada (D1), segunda derivada (D2), remoção contínua (CR), variação normal padrão (SNV), Savitzky-Golay (SG) e Savitzky-Golay com primeira derivada (SG-D1), respectivamente. Na escala observada é evidente que, ao comparar os espectros brutos sem transformação (Figura 10) com o pré processamento variação normal padrão (SNV) (Figura 11 D), minimizou-se uma parte do ruído contido na informação espectral centrando o conjunto de dados em torno de zero, devido a isso sua escala espectral tornou-se diferente. Confirma-se que os dados sem transformação e pré processados por SNV, possuem o mesmo deslocamento e posicionamento das bandas nas regiões do espectro.

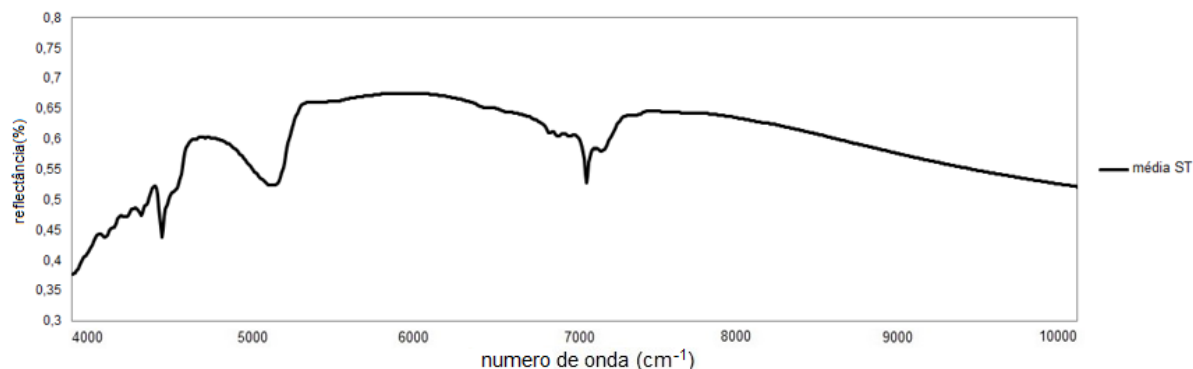


Figura 10 - Reflectância dos espectros na região NIR (4000 cm^{-1} – 10000cm^{-1}) sem transformação (ST)

Neste trabalho, com objetivo de remover informações físicas nos espectros, foram avaliados alguns métodos de pré-processamento de dados e corrigida a linha de base dos espectros. Na região acima de 7500 cm^{-1} observou-se que não havia informações relevantes sobre as amostras e apresentava somente ruído. O método da primeira derivada de Savitzky-Golay foi eficiente devido a sua capacidade de deslocar o eixo dos espectros para o zero e remover deslocamentos verticais de linha de base. Estes são problemas reais ao medir os espectros, produzindo variações indesejadas nas medidas de reflectância conforme demonstrado na Figura 11 (A) e (F). Ao utilizar a segunda derivada, eliminou-se a inclinação dos espectros, conforme visto na Figura 11 (B), porém todas as transformações mantiveram a informação da área dos picos conservadas.

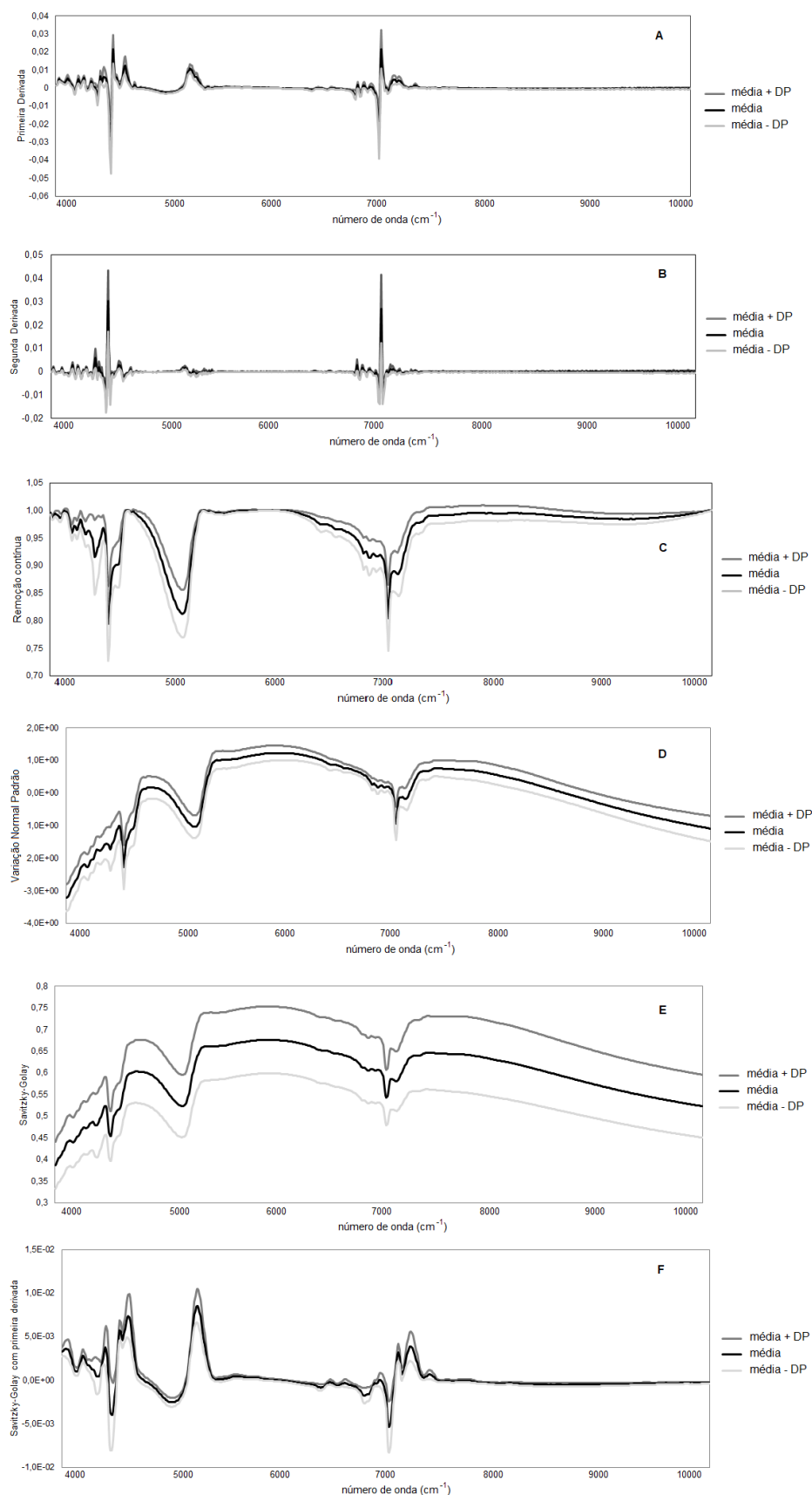


Figura 11 - Reflectância dos espectros na região NIR pré processados por (A) primeira derivada, (B) segunda derivada, (C) remoção contínua, (D) variação normal padrão, (E) Savitzky-Golay e (F) Savitzky-Golay com primeira derivada

4.3.1 Calibração Multivariada - PLS

As Tabelas 4 e 5 apresentam as estatísticas dos modelos feitos por regressão mínimos quadrados parciais e validação para os teores de COT e NT, respectivamente. Para modelagem de dados de espectroscopia NIR de amostras de solo ocorre maiores valores de componentes principais, raiz quadrada do erro médio, devido a grande diferença e complexidade deste tipo de amostra (SOUZA, 2014). O número de componentes principais utilizado no modelo é determinado pela porcentagem de variância explicada. Assim, seleciona-se um número de componentes de tal maneira que a maior porcentagem da variação presente no conjunto de dados originais seja capturada. Este tipo de comportamento se mostrou apenas para análise do teor de COT, já que para NT, os valores foram baixos devido a unidade e magnitude dos mesmos nas amostras. Avaliando os valores da raiz quadrada do erro médio entre os dados de calibração e entre os dados validação, percebe-se que são estatisticamente semelhantes para os dois atributos em análise, porém os erros quadráticos médios de validação foram menores que os erros quadráticos médios de calibração para todas as características analisadas, exceto para variação normal padrão em NT. Por tratar-se de erros relativos médios, a avaliação dentro de cada intervalo fez-se necessária para observar erros muito discrepantes aos valores médios encontrados e até identificar possíveis tendências de erros em determinadas faixas de concentração dos analitos.

Tabela 4: Avaliação dos modelos de calibração e validação obtidos por mínimos quadrados parciais na determinação de COT.

Pré-Trat.	CP	CV (%)		RMSE		RPIQ		Linearidade Resíduos - bias	
		Cal	Val	Cal	Val	Cal	Val	Cal	Val
ST	22	38,35	37,85	3,56	2,77	1,92	2,49	-0,13	0,43
MSC	20	34,57	36,64	3,79	2,87	1,67	2,10	-0,41	0,13
D1	10	37,76	35,49	3,85	3,20	1,75	1,90	-0,10	0,36
D2	4	37,46	33,61	3,80	3,24	1,88	2,13	-0,01	0,13
CR	14	37,00	34,93	3,94	3,01	1,70	2,08	-0,17	0,51
SNV	21	38,06	35,52	3,41	2,76	2,02	2,25	-0,12	0,47
SG	22	37,93	37,19	3,69	2,83	1,82	2,43	-0,13	0,39
SG+ D1	11	37,12	35,36	3,93	3,10	1,60	1,97	-0,09	0,38

Legenda: ST – Sem tratamento; MSC – correção multiplicativa de sinal; D1 – 1ª derivada; D2 – 2ª derivada; CR – remoção contínua; SNV – variação normal padrão; SG – Savitzky-Golay; CP – componentes principais; RMSE – raiz quadrada do erro médio; RPIQ – relação entre o desempenho e a distância interquartil.

Tabela 5: Avaliação dos modelos de calibração e validação obtidos por mínimos quadrados parciais na determinação de NT.

Pré-Trat.	CP	CV (%)		RMSE		RPIQ		Linearidade Resíduos - bias	
		Cal	Val	Cal	Val	Cal	Val	Cal	Val
ST	19	35,72	32,16	0,24	0,20	1,73	2,43	-0,01	-0,002
MSC	18	31,21	27,99	0,25	0,22	1,62	1,71	0,00	0,001
D1	9	27,02	25,12	0,24	0,21	1,75	1,96	-0,02	0,017
D2	10	31,43	26,87	0,23	0,21	1,73	1,90	-0,01	-0,020
CR	14	31,57	27,07	0,24	0,20	1,75	2,36	-0,01	0,001
SNV	19	30,89	27,91	0,32	0,33	1,04	0,73	0,00	-0,017
SG	18	23,71	22,88	0,24	0,20	1,66	2,61	-0,01	0,025
SG+ D1	11	31,04	28,06	0,25	0,22	1,60	2,21	-0,01	-0,011

Legenda: ST – Sem tratamento; MSC – correção multiplicativa de sinal; D1 – 1ª derivada; D2 – 2ª derivada; CR – remoção contínua; SNV – variação normal padrão; SG – Savitzky-Golay; CP – componentes principais; RMSE – raiz quadrada do erro médio; RPIQ – relação entre o desempenho e a distância interquartil.

Para avaliar a qualidade dos modelos sem tratamentos para validação, e a variação normal padrão para calibração, é recomendado a utilização da relação entre o desempenho e a distância interquartil, devido a disposição dos teores de COT nas amostras de solo utilizadas neste trabalho. Para NT, o modelo primeira derivada e remoção contínua apresentou maior valor para calibração e o Savitzky-Golay para validação. Sendo assim, estes apresentaram melhor a capacidade preditiva dos modelos de COT e NT. A aplicação do teste bias, foi aplicado para erros sistemáticos que afetam a exatidão das medidas, sendo associados a todas as componentes do erro que não são aleatórias. Este valor pode ser positivo ou negativo. Para calibração do COT os resultados foram negativos e indicam que as previsões do NIR variam significativamente a partir dos valores experimentais seguindo valores maiores, o que percebe-se no processamento com correção multiplicativa de sinal. Em contrapartida, para validação os resultados foram positivos, indicando que o modelo estava superestimado, o que mais se aplica para o processamento com remoção contínua, enquanto que o modelo com correção multiplicativa de sinal e segunda derivada foram os menos estimados em relação ao bias. Ao atributo NT, ambos (calibração e validação) apresentaram valores negativos, variando assim as previsões do NIR a este.

Na etapa de pré-processamento dos dados espectrais foram também avaliados os 8 métodos empregados em espectroscopia NIR, e outra forma de apresentar os resultados é em forma de gráficos com os valores obtidos em laboratórios e o previsto

pelo NIR, conforme observado nas Figuras 12, 13, 14 e 15. Essas figuras mostram a distribuição dos pontos de calibração e validação dos melhores modelos para análise de COT e NT. Para o atributo COT existe menor dispersão, o que confirma melhor predição dos modelos e o que significa que para um mesmo valor de teor de COT medido pelo método de referência adotado, a metodologia NIR prevê valores próximos, o que causa um menor espalhamento dos pontos em torno da linha de tendência. O modelo para os dados processados com segunda derivada proporcionou maior coeficiente de determinação tanto para calibração (0,866) quanto para validação (0,889), sendo assim o modelo está mais correlacionado. Em contrapartida, o modelo aplicado com remoção contínua apresentou o menor valor para calibração (0,643).

Para NT, percebeu-se comportamento contrário ao COT, pois a dispersão dos pontos mostraram-se maiores, afirmando uma menor linearidade dos modelos e o que significa que para um mesmo valor de teor de NT medido pelo método de referência adotado, a metodologia NIR prevê valores mais afastados, o que causa maior espalhamento dos pontos em torno da linha de tendência. O modelo para os dados processados com segunda derivada proporcionou pior coeficiente de determinação tanto para calibração quanto para validação, sendo assim o modelo não é linear. Em contrapartida, o modelo aplicado com variação normal padrão apresentou o melhor valor para calibração (0,677) e para validação o maior valor apresentado foi o Savitzky-Golay (0,656).

Desta forma, os modelos construídos podem ser considerados lineares, e classificados para calibrar e validar em melhor posição para o teor COT o modelo aplicado segunda derivada. O modelo com melhor desempenho para o teor de NT para calibração foi o variação normal padrão e para validar o modelo, o tratamento com Savitzky-Golay. Ambos que apresentaram o melhor coeficiente de determinação, uma vez que seus resíduos apresentaram comportamento aleatório, e a magnitude dos erros foi da mesma ordem (Tabelas 5 e 6). Devido a forte associação entre os valores mensurados e preditos pelo modelo, indica-se a possibilidade de uso da técnica NIRS para estimar o teor de COT. Enquanto às bandas de teores de NT, deve-se melhor verificar os erros e aplicação de outros tratamentos na análise, enquanto que seu resultado pode ter sido menos acurado devido a sobreposição às bandas de C.

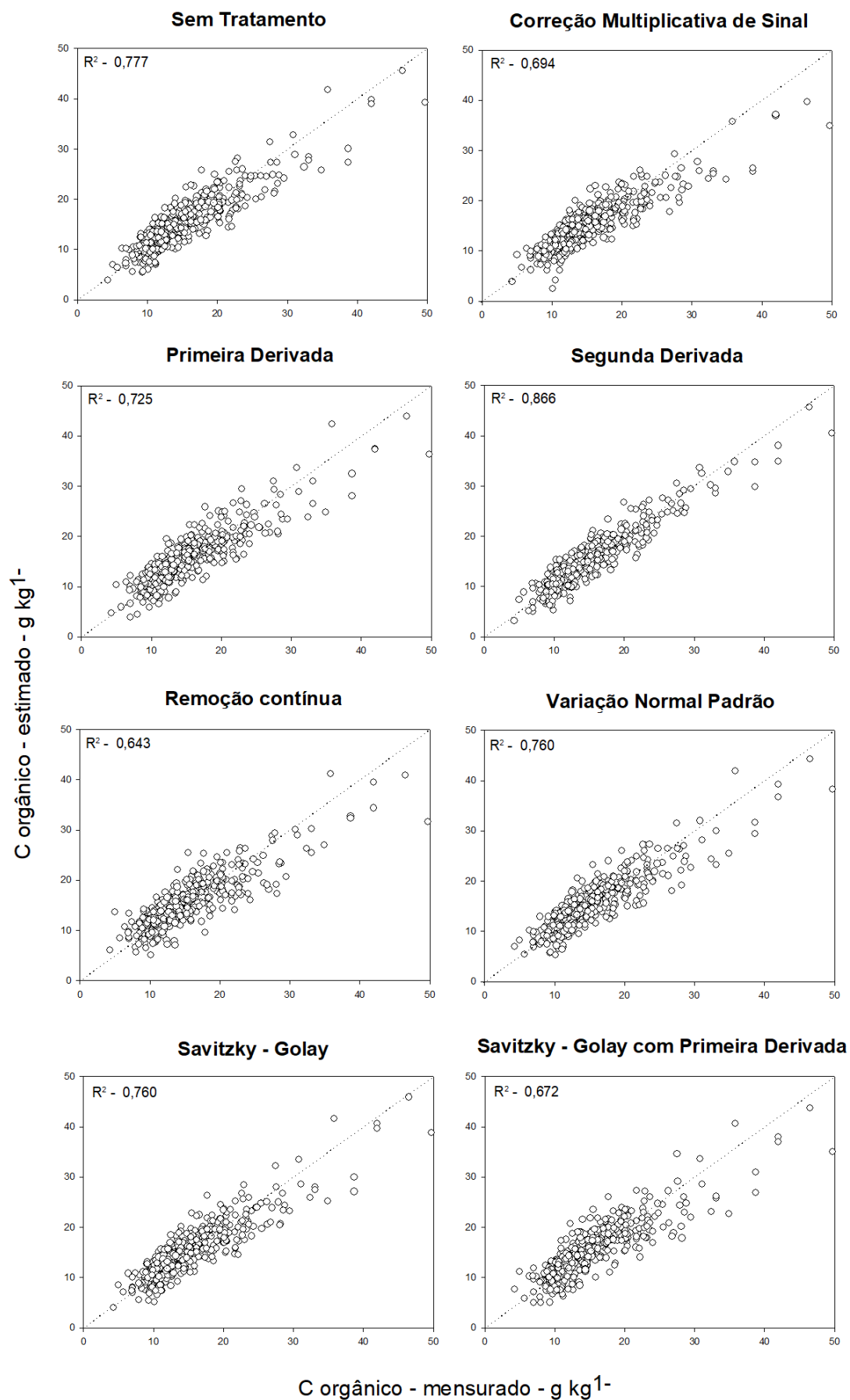


Figura 12 – Desempenho do modelo para calibração dos dados de COT.

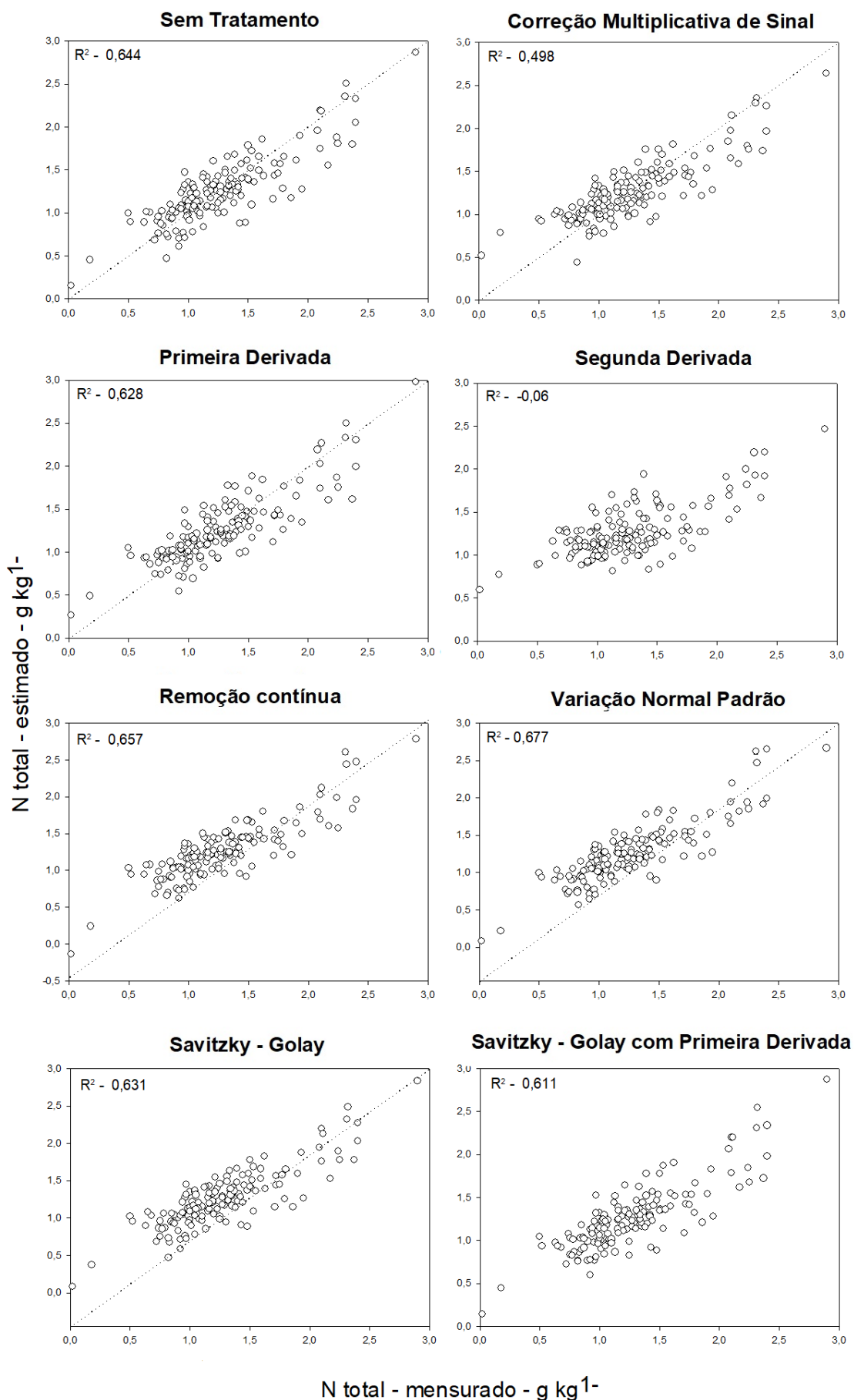


Figura 13 – Desempenho do modelo para calibração dos dados de NT.

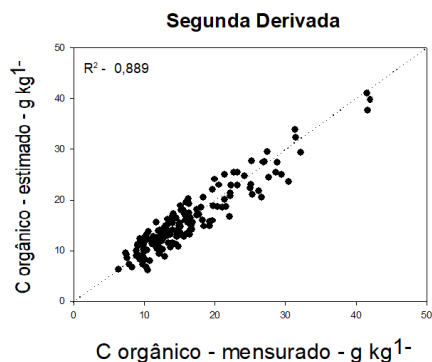


Figura 14 – Desempenho do modelo para validação dos dados de COT.

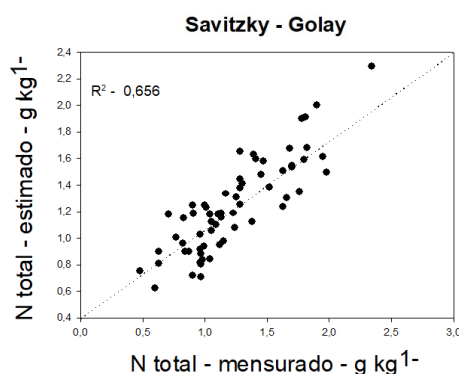


Figura 15 – Desempenho do modelo para validação dos dados de NT.

Neste trabalho, o método com melhor desempenho para calibração e validação das amostras em análise de COT foi o aplicação com segunda derivada ($R^2 = 0,866$ e $R^2 = 0,889$, respectivamente). Já para análise de NT, o melhor desempenho para calibração foi o método com variação normal padrão ($R^2 = 0,677$), e para validação foi o aplicação com Savitzky-Golay ($R^2 = 0,656$). As figuras 16 e 17 mostram os gráficos dos coeficientes da equação em cada número de onda. Percebe-se na figura 16, para aplicação com segunda derivada, que o melhor desempenho do coeficiente da equação para análise de COT, foi no número de onda entre 5000 cm^{-1} e 5500 cm^{-1} . A análise de NT, com aplicação variação normal padrão, teve comportamento com melhor desempenho próximo aos 4500 cm^{-1} , como apresenta a figura 17. Ou seja, para análise de COT e NT corrobora-se o melhor comportamento na faixa espectral entre 4000 e 5000 cm^{-1} .

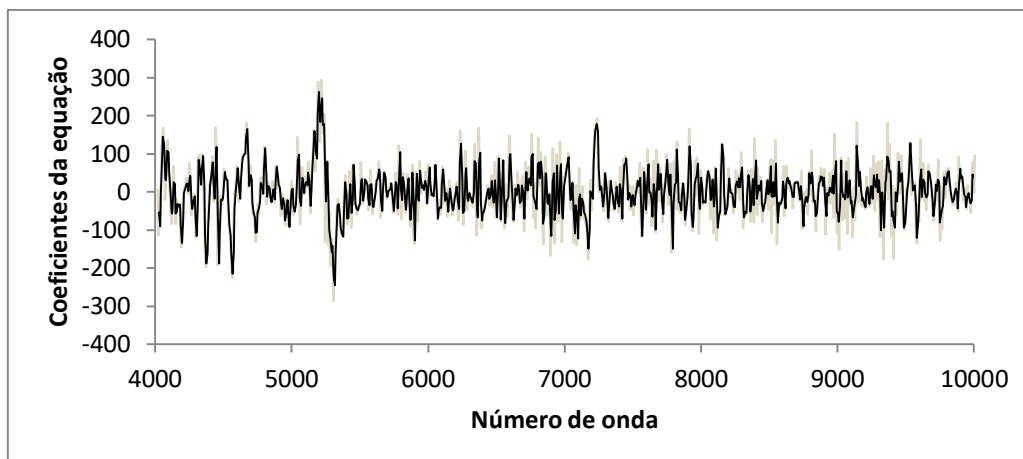


Figura 16 - Coeficientes da equação em cada número de onda para análise de COT.

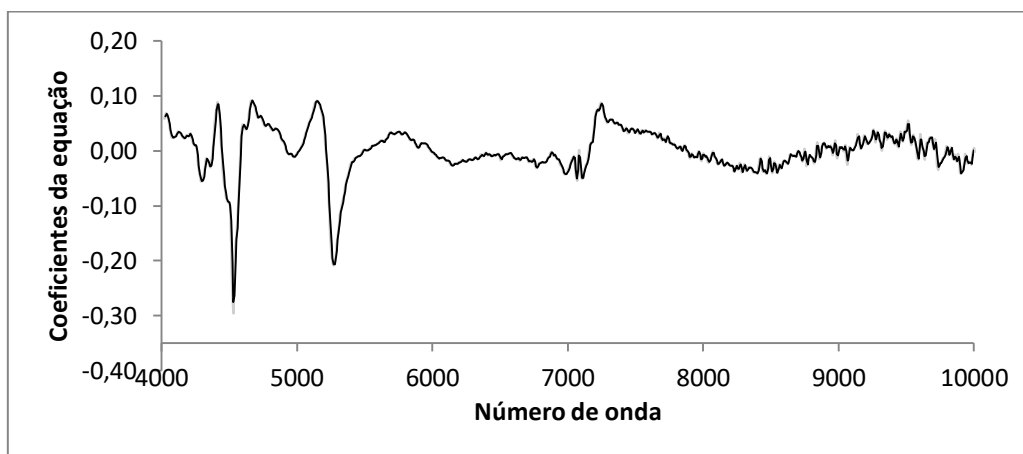


Figura 17 - Coeficientes da equação em cada número de onda para análise de NT.

De acordo com a tabela 1, que diz respeito aos trabalhos relacionados à técnica NIR e análise de solos, em CO: SOUZA, 2014 obteve R^2 de 0,83, SATO, 2013 obteve R^2 entre 0,80 e 0,88, FELIX, et al. 2016, $R^2 = 0,86$, SEGNINI, et al. 2014, foi o que melhor apresentou $R^2 = 0,99$, KUSUMO, SUKARTONO, BUSTAN, 2018, obtiveram $R^2 = 0,76$. Para o teor de NT, FELIX, et al. 2016 obteve $R^2 = 0,97$, valor alto pois utilizou um único tipo de solo, o basalto, sendo assim sua homogeneidade foi maior fazendo com que como consequência obtenha um coeficiente alto e MARCHÃO, R. L., BECQUER, T., BRUNET, D, 2011, $R^2 = 0,11$ para calibração e 0,58 para validação.

Em comparativo aos trabalhos referenciados, este trabalho obteve valores satisfatórios e com bons ajustes para calibração e validação tanto para análise de COT, quanto para análise de NT.

5 CONCLUSÕES

O presente estudo mostrou que foi possível determinar os teores de COT e NT para amostras de solo de maneira rápida e com confiabilidade a partir da espectroscopia de infravermelho próximo.

O pré-tratamento dos dados espectrais das amostras contribuíram para uma melhor predição das propriedades estudadas. Neste estudo o pré-tratamento que possibilitou a construção de um modelo mais robusto e com uma capacidade preditiva maior que os demais foi a segunda derivada para o teor de Carbono orgânico total e variação normal padrão para o teor de Nitrogênio total. Os modelos utilizados para validação dos nutrientes do solo, requerem um banco de dados com maior amplitude em relação aos teores e maior quantidade de amostras, para assegurar a representatividade do espaço amostral e confiança e precisão dos modelos de calibração e validação. O teor de Carbono orgânico total utilizado neste trabalho variou de 4,3 a 52,80 g Kg⁻¹, enquanto que o teor de Nitrogênio total variou de 0,02 a 2,90 gKg⁻¹, o que percebe-se a importância de uma maior amplitude em relação a esses dados obtidos, principalmente relacionado ao teor de Nitrogênio total.

Os modelos de predição de C orgânico total foram considerados satisfatórios na utilização da técnica NIRS, aliada a quimiometria, tornando-se uma técnica alternativa para análise dos teores de Carbono orgânico e Nitrogênio total do solo.

Através dos parâmetros de predição, este trabalho evidenciou que a utilização de pré-tratamentos espectrais distintos, seleção do número de componentes principais e utilização de região espectral específica na construção dos modelos, influenciam na obtenção de um modelo mais confiável para quantificação de COT e NT em solos.

6 REFERÊNCIAS

AMORIM, H. V. **Manual de métodos analíticos para o controle da produção de álcool e açúcar**. 194p. 2. ed. Piracicaba: ESALQ-USP, 1996

BARBARA. **Espectroscopia no infravermelho**. 2013 Disponível em <<http://www.ufjf.br/quimicaead/files/2013/05/ESPECTROSCOPIA-NO-INFRAVERMELHO-PARTE1.pdf>>.

BELTRAME, K. K. **Avaliação dos métodos Walkley & Black e CHN como métodos de referência para calibração multivariada na determinação de Carbono orgânico em solos brasileiros**. 2014. 48f. Trabalho de Conclusão de Curso (TCC) – Curso de licenciatura em química, Universidade Tecnológica Federal do Paraná – UTFPR. Campo Mourão - Paraná. 2014.

BENACI, V. **Avaliação de métodos de análise para Carbono orgânico em amostras de Interesse agrônomo**. 2010. 63f. Dissertação (mestrado em Agricultura Tropical e Subtropical) - curso de Pós-Graduação em Agricultura Tropical E Subtropical, Instituto agrônomo. Campinas – SP. 2010.

BOSQUILIA, R. W. D. **Comportamento espectral de alvos: Solos, águas e outros**. Disciplina de Sensoriamento Remoto. Universidade Tecnológica Federal do Paraná – UTFPR. Paraná. [s.d].

CHANG, C.W. et al. **Near-Infrared Reflectance Spectroscopy–Principal Components Regression Analyses of Soil Properties**. Published in Soil Sci. Soc. Am. J. 65:480–490. 2001.

Chen, H. et al. **The combined Optimization of Savitzky-Golay Smoothing and Multiplicative Scatter Correction for FT-NIR PLS Models**. Hindawi publishing corporation. Volume 2013. Article ID 642190. 9p. China. 2012.

CHRÉTIEN, J. R. **Boosting Chemometrics in Europe**. Chemom. Intell. Lab. Syst. 1-2 p. 2003.

CIOTTA, M. N. et al. Soil organic matter and cation exchange capacity increase in a low activity clay soil under no-tillage system. **Ciência Rural**, v. 33336, n. 6, p. 1161–1164, 2003.

COSTA. **Espectroscopia no Infravermelho 3**. PUC RIO – certificação digital nº 0114349/CA. 2003. Disponível em:<https://www.maxwell.vrac.pucrio.br/4432/4432_4.PDF>

COSTA, L. R. **Aplicação da espectroscopia no infravermelho próximo para avaliação do carvão vegetal**. 2017. 62f. Dissertação (mestrado em processamento e utilização da madeira – curso de Pós – Graduação em Ciência e tecnologia da Madeira, Universidade Federal de Lavras – UFLA. Lavras - Minas Gerais. 2017.

DALMOLIN, R. S. D; GONÇALVES, C. N; KLAMT, E; DICK, D. P. Relações entre os constituintes do solo e seu comportamento espectral. **Ciência Rural**, Santa Maria, v. 35, n. 2, p. 481- 489, 2005.

DEHANOV, N. F. **Técnicas de calibração espectroscópica baseadas na estimativa do ruído espectral e do sinal de resposta aplicadas a espectros de infravermelhos médios (MIR) de amostras de culturas de células estaminais**. 2011. 58f. Dissertação (mestrado em Engenharia Farmacêutica) – curso de Pós-Graduação em Engenharia Farmaceutica, Universidade técnica de Lisboa – Instituto Superior técnico. Lisboa. 2011.

DIAS, C. **Espectroscopia NIR aplicada à análise de solos foi tema de debate em workshop nacional**. Embrapa Solos. 2014. Disponível em: <http://www.diadecampo.com.br/zpublisher/materias/Materia.asp?id=31095&secao=A_grotemas>. Acesso em: 29 de Janeiro de 2019

DONATO E.M.; et al. **Espectrofotometria derivada: uma contribuição prática para o desenvolvimento de métodos**. Revista de Ciências Farmacêuticas Básica e aplicada, 31(2): 125-130p. 2010.

DONOSO, J. P. **Datas e personagens na história da espectroscopia**. Universidade de São Paulo, Instituto de Física de São Carlos - IFSC. [s.d]

DOTTO, A. C; et al. **A systematic study on the application of scatter-corrective and spectralderivative preprocessing for multivariate prediction of soil organic carbono by Vis-NIR spectra**. Geoderma, 314: 262–274p. 2018.

ELISABETH, M.; CLAESSEN, C. **1997_Embrapa_Analise-Solo**. [s.l: s.n.].

FELIX, J. C. et al. **Predição de fósforo, carbono e nitrogênio em solos de basalto, por meio de espectroscopia NIR**. Pesq. agropec. bras., Brasília, v.51, n.9, p.1405-1416, 2016.

FERREIRA, M. M. C. **Quimiometria: Conceitos, métodos e aplicações**. Campinas, SP: Editora da Unicamp, 2015.

FIDÊNCIO, P. H. **Análise de Solos por Espectroscopia no Infravermelho Próximo e Aplicação de Métodos Quimiométricos**. 2001. 149f. Tese (doutorado em Química analítica) – Curso de Pós-Graduação em ciências, Universidade Estadual de Campinas- UNICAMP- Instituto de Química. CAMPINAS, SÃO PAULO. 2001.

FRANCESCHINI, M. H. D. et al. **Abordagens semiquantitativa e quantitativa na avaliação da textura do solo por espectroscopia de reflectância bidirecional no VIS-NIR-SWIR**. Pesquisa Agropecuária Brasileira, v. 48, n. 12, p. 1569–1582, 2013.

GUIMARÃES, P. T. G.; ALVAREZ, V. H.; RIBEIRO, A. C. **Amostragem do Solo. Recomendações Para o Uso de Corretivos e Fertilizantes em Minas Gerais - 5º Aproximação**, p. 13–20, 1999.

ILHARCO, L. M. **Espectroscopia de Infravermelho: Uma técnica Antiga, sempre**

atual. Sociedade portuguesa de Química. 1998.

JÚNIOR, C. C. **Matéria orgânica, capacidade de troca catiônica e acidez potencial no solo com dezoito cultivares de cana-de-açúcar**. 2011. 134F. Tese (doutorado em Agronomia (produção vegetal)) – Universidade Estadual Paulista “Julio de Mesquita Filho” - Faculdade de Ciências Agrárias e Veterinárias. Jaboticabal, SP. 2011.

Kusumo B H, Sukartono and Bustan. **The rapid measurement of soil carbon stock using near-infrared technology**. In IOP Conference Series: Earth and Environmental Science 129. 2018

LEITE, L. F. C. EMBRAPA. **Matéria Orgânica do solo**. TERESINA: EMBRAPA Meio-Norte. 1ª ed. 31p. 2004.

LÓPEZ, L. R. **Pedologia quantitativa: espectrometria VIS-NIR-SWIR e mapeamento digital de solos**. 2009. 171F. Dissertação (mestrado em agronomia) – Universidade de São Paulo – Escola Superior de Agricultura “Luiz de Queiroz”. Piracicaba, SP. 2009.

MARCHÃO, R. L., BECQUER, T., BRUNET, D. **Predição dos teores de Carbono e Nitrogênio do solo utilizando espectroscopia de Infravermelho próximo**. Boletim de pesquisa e desenvolvimento . Empresa Brasileira de Pesquisa Agropecuária – Embrapa Cerrados. Planaltina, DF. 2011.

MENDONÇA, E. D., & MATOS, E. D. **Matéria Orgânica do Solo: Métodos de Análises**. 2ª ed. Revista e Atualizada. 221p. Viçosa, MG: UFV – Gefert. 2017.

MENEGHIN, M. F. S. et al. **Avaliação da disponibilidade de nitrogênio no solo para o trigo em latossolo vermelho do distrito federal**. Revista Brasileira de ciência do Solo. *SciELO - Scientific Electronic Library Online*. 32: pag.1941-1948. 2008. Disponível em: <<http://www.scielo.br/pdf/rbcs/v32n5/15.pdf>>.

MEYER, J. H. Use of NIR in the South African sugar industry with reference to soil fertility management. **South African Sugar Association Experiment Station**, KwaZulu-Natal. 1999.

MONTEIRO, A. da R. D. **Estudo de Aspectos Estatísticos da Espectroscopia na Região do Infravermelho Próximo**. 2011. 183f. Dissertação (mestrado em Engenharia Química) – Programa de Pós – Graduação e Pesquisa e Engenharia – Universidade Federal do Rio de Janeiro – UFRJ. Rio de Janeiro, RJ. 2011.

MORÓN, A. & COZZOLINO, D. **Exploring the use of near infrared reflectance spectroscopy to study physical properties and microelements in soils**. Journal of Near Infrared Spectroscopy, v.11, p.145-154, 2003.

NISHIDA, E. N. **Propriedades da Filtragem de Savitzky-Golay Aplicadas na Identificação de Complexos QRS em Sinais de Eletrocardiograma**. 2017. 53f. Dissertação (Mestrado em Ciência em Engenharia elétrica) – Programa de Pós – Graduação em engenharia Elétrica – Universidade Federal de Itajubá. Itajubá, MG.

2017.

OLIVEIRA, J. F. de. **Resposta espectral Vis-NIR dos microagregados de latossolos do planalto central**. 2015. 191f. Tese (Doutorado em Agronomia) – Programa de Pós Graduação em Agronomia – Universidade Estadual de Londrina – Centro de Ciências Agrárias. Londrina – PR. 2015.

PAN, T., et al. **Optimal partner wavelength combination method with application to near-infrared spectroscopic analysis**. *Elsevier*, 7p. 2016.

PAULO, U. D. S.; JOSÉ, P.; DONOSO, P. **Datas e personagens na História da Termodinâmica**. [s.d.].

RANGEL, O. J. P. et al. **Carbono orgânico e nitrogênio total do solo e suas relações com os espaçamentos de plantio de cafeeiro**. *Ci solo*. 32:2051-2059, 2008.

ROMÃO, R. L. **Carbono orgânico em função do uso do solo**. 2012. 47f. Dissertação (Mestrado em Agronomia – Ciência do solo) - Programa de pós graduação em Agronomia – Ciência do solo, Universidade Estadual Paulista – UNESP, JABOTICABAL, SP. 2012

SANTOS, A. P. dos. **Espectroscopia de Infravermelho próximo em análise de solos e plantas**. 2011. 73f. Dissertação (mestrado em Agronomia) – Programa de pós graduação em agronomia, Universidade Federal de Uberlândia. Instituto de ciências agrárias. Uberlândia - MG. 2011.

SANTOS, G.; PEREIRA, A.; KORNDÖRFER, G. **Use of analysis system by near infrared (NIR) for the analysis of organic matter and clay fraction in soils and leaf levels of silicon and nitrogen in sugar cane**. *Bioscience Journal*, Uberlândia, v. 26, n. 1, p. 100-108, Jan./Feb. 2010.

SATO, J. H. **MÉTODOS PARA DETERMINAÇÃO DO CARBONO ORGÂNICO EM SOLOS DO CERRADO**. 2013. 90F. Dissertação (mestrado em Agronomia) – Programa de pós graduação em agronomia, Universidade de Brasília. Faculdade de Agronomia e Medicina Veterinária. Brasília – DF. 2013.

SEGNINI, A. et al, **DETERMINAÇÃO DE PROPRIEDADES DO SOLO UTILIZANDO ESPECTROSCOPIA DE INFRAVERMELHO PRÓXIMO**. Simpósio Nacional de Instrumentação Agropecuária, São Carlos, Nov. 2014.

SERRAT, B.M.; KRIEGER, K.I. & MOTTA, A.C.V. **Considerações sobre a interpretação de análise de solos (com exemplos)**. *Recomendações de manejo do solo: Aspectos tecnológicos e metodológicos*. p.125- 143. Universidade Federal do Paraná, Curitiba - PR. 2006.

SHEPHERD, K. D.; WALSH, M. G. **Development of Reflectance Spectral Libraries for Characterization of Soil Properties**. *Soil Science Society of America Journal*, v. 66, n. 3, p. 988, 2002.

SILVEIRA, A. J. de A. **Espectroscopia na região Infravermelho**. Cap 8-9. 24p. 2009. Disponível em: < <https://www.doraci.com.br/downloads/engenharia/espectroscopia-infravermelho.pdf> >

SOUZA, A. M. **Avaliação do potencial da espectroscopia no infravermelho próximo como método de rotina para a determinação de carbono orgânico do solo**. 2014. 159f. Tese (doutorado em ciências) – Programa de Pós-Graduação em Ciências, Universidade Estadual de Campinas – UNICAMP, Instituto de Química. CAMPINAS, SP. 2014.

SOUZA, A. M.; POPPI, R. J. **Experimento didático de quimiometria para análise exploratória de óleos vegetais comestíveis por espectroscopia no infravermelho médio e análise de componentes principais: um tutorial, parte I**. *Quim. Nova*, Vol. 35, No. 1, 223-229f, 2012.

STATSOFT COMPANY HISTORY. Statsoft, **Statistica, version 8.0**. 2008. Disponível em: <<http://www.statsoft.com/company/history/>>.

SUNOJ, S.; IGATHINATHANE. C; VISVANATHAN, R. **Nondestructive determination of cocoa bean quality using FT-NIR Spectroscopy**. *Computers and Electronics in Agriculture* 124, 234–242p. 2016. Disponível em: <<http://dx.doi.org/10.1016/j.compag.2016.04.012>>.

SYSTAT SOFTWARE Inc. – SSI. **Sigmaplot for Windows, version 10**. 2006. Disponível em: <<http://www.systat.com/products/sigmaplot/>>.

TEÓFILO, R.F., **Chemometric methods in the electrochemical studies of phenols on boron-doped diamond films**, Universidade Estadual de Campinas, Campinas, SP. 2007.

TEÓFILO, R.F., **Métodos Quimiométricos: Uma Visão Geral - Conceitos básicos de quimiometria**, Universidade Federal de Viçosa, Viçosa, MG. Vol. 1. 2013.

VISCARRA ROSSEL, R.A., CATTLE, S.R., ORTEGA, A. & FOUAD Y. **Geoderma**, 150, p.253-266, 2009.

WALKLEY, A.; BLACK, I. A. **An examination of the Degtjareff method for determining soil organic matter and a proposed modification of the chromic acid titration method**. *Soil Science*, Baltimore, v. 37, p. 29-38, 1934.

WALKLEY, A. & BLACK, C. A. **Determination of Organic carbon in soil**. In: JUO, A.S.R., ed. *Selected methods for soil and plant analysis*. 2 ed. Ibadan, International Institute of Tropical Agriculture, 1978. P. 4-5.

Yu, X., Liu, Q., YebaoWang, Liu, X., & Liu, X. **Evaluation of MLSR and PLSR for estimating soil element contents using visible/near-infrared spectroscopy in apple orchards on the Jiaodong peninsula**. *Elsevier*, 340-349p. 2015.